欠測値のあるソフトウェアプロジェクトデータに基づく信頼性解析

2014 年度 修士論文

欠測値のあるソフトウェアプロジェクトデータに基づく信頼性解析

法政大学大学院理工学研究科
システム工学専攻

13R6214　森田　貴之

Takayuki MORITA

指導教員　木村　光宏 教授

概　要

ソフトウェアリリース後の不具合有無をその開発途中で予測することができれば，ソフトウェア信頼性に大きく貢献する．この事実を受けて多変量解析を用いてソフトウェアの信頼性予測を行うことにしたが，予測を行うために必要なソフトウェアプロジェクトデータには多くの欠測値が含んでいることが分かった．そこで本研究では完全データとそれに類似した不完全データを元データから構築し，不完全データの欠測値を代入することで擬似完全データを作成した．完全データと擬似完全データを対象データにとして信頼性予測を行った結果，予測精度にほとんど差が生じない方法を発見することができた．予測精度に差が生じないことから，擬似完全データに対して完全データとほぼ遜色のない信頼性予測ができる可能性を示した．

Abstract

Software reliability is more enhanced if a software development project manager can predict whether the number of failures after the software release is zero or not, during the software development process. In response to this fact, we started this investigation which predicts the software reliability by using the multivariate analysis approach. Unfortunately, the objective data set we obtained has a lot of missing values. Therefore in this study, we first prepared two data sets. One is the complete data set extracted from the raw data set, and the other was created as a virtual complete data set by means of the data imputation method so as to be similar to the complete data set. As a result of the software reliability prediction, we have found that our imputation method is comparable in terms of the prediction accuracy between two data sets via the discriminant analysis. Therefore it is shown that the data sets with our missing value treatment method can be used for software reliability prediction as well as the complete data sets.

目次

表目次

図目次

1　はじめに

この章では本研究の背景とそれに関連した先行研究，背景と先行研究を受けた研究目的の説明をする．章ごとに目的を達成するために行った研究についても本論文の構成として記述した．

1.1　研究背景と目的

本研究で取り扱うソフトウェアとは，「情報処理システムのプログラム，手続き，規則及び関連文書の全体又は一部分」のことである [1]．ソフトウェアはプログラムという命令から成り立っており，1949 年にウィルクスがプログラムで動くコンピュータを開発してから現在までにソフトウェアは我々にとって非常に身近なものとなっている．もともとソフトウェアはコンピュータの計算を自動化させるツールにすぎなかった．しかし，現在では E-mail に代表されるコミュニケーションツールや銀行のオンラインシステムにとどまらず，飛行機の操縦システムなどの我々の命を預かる役目も担うようになった．

ソフトウェアプロジェクトとはソフトウェア開発の計画のことを指す．開発計画はおおまかに計画と開発，運用・保守の 3 つに分類することができる．これらのサイクルをより細かく分類することで実際の開発工程と決めることができるが，開発プロジェクトによって分類が若干異なる場合がある．本研究では図 1 のように取り決めた [2]．

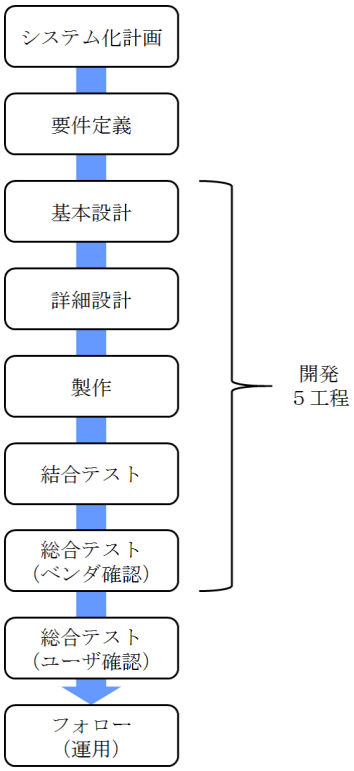

図 1　ソフトウェア開発工程．

システム化計画

この工程ではソフトウェアが実現すべき機能などの確認と整理を行う．これによって課題を定義し，業務機能をモデル化する．得られたモデルを用いてソフトウェア品質の基本方式の明確化，プロジェクトの目標設定，全体開発スケジュールの作成などを行った後に取り決め内容をシステム化計画として文書化し，ステークホルダの合意を得る．

要件定義

ソフトウェアに関する実現可能性を検証する．その後にシステム設計が可能な技術要件に変換し，ソフトウェア要件を文書化する．また，設定した基準を考慮してソフトウェア要件を評価し，文書化する．

基本設計

ソフトウェアを構成する品目とそれに伴って必要となる手作業を明確にし，これらを作業に割り振ったソフトウェア要件を文書化する．また，ソフトウェアを構成する品目に対する要件をソフトウェア方式に変換する．ソフトウェアコンポーネントやデータベースの設計，ソフトウェア結合のためのテスト要求事項の明確化など開発に必要な基本的な設計を始める．

詳細設計

ソフトウェア品目の各ソフトウェアコンポーネントに対して詳細設計を行う．ソフトウェアコンポーネントはコーディング，コンパイルおよびテストを実施するユニットレベルに詳細化する．また，ソフトウェアインターフェイスとデータベースの詳細設計を行い，必要に応じて利用者文書を更新する．加えて，ユニットテストや結合テストのためのテスト要求事項および予定を定義する．

製作

ソフトウェアユニットおよびデータベースを開発する．加えて，それらに対するテスト手順およびデータを設定する．その後，テストを実施して要求事項を満たしているかどうかを確認する．これらに基づいて必要があれば文書を更新する．

結合テスト

ソフトウェアユニットとソフトウェアコンポーネントを結合して，ソフトウェア品目にするための計画を作成し，ソフトウェア品目を完成させる．その後，結合とテストを行う．要件を満たしているかどうかの最終確認を実施可能状態にする．

総合テスト（ベンダ確認）

開発者が指定された要求事項に従って確認テストと評価を行う．

総合テスト（ユーザ確認）

実環境にソフトウェアを導入するための計画を作成した後にソフトウェアを導入する．また，開発者は取得者によるソフトウェアの受け入れレビューとテストを支援する．

フォロー（運用）

ソフトウェアの運用および利用者に対する運用支援を行う．運用者はプロセスを管理するために具体化した管理プロセスに従って運用プロセスの基盤となる環境を構築する．

システム化計画から要件定義までが計画にあたり，基本設計から総合テスト（ベンダ確認）までが開発，総合テスト（ユーザ確認）からフォローまでが運用・保守にあたる [3]．また，基本設計から総合テスト（ベンダ確認）までを開発５工程と呼ぶ．このようにソフトウェア開発は細かく体系化されている．この理由は信頼性の高いソフトウェアを作成するためである．この開発工程を通して信頼性が高いソフトウェアをつくることを目標に進める．本研究が定める信頼性とはソフトウェアにバグを含むかどうかである．信頼性という観点で開発工程に着目するとシステム化計画と要件定義では顧客の要求という側面を中心にソフトウェアについての議論が交わされる．その際に次の段階でバグを含まないようなソフトウェア設計と開発の方法についても決定される．基本設計，詳細設計，製作では前段階で取り決められた方法にそって開発が進むため，人的エラーによって生じるバグの混入に注視して進められる．結合テストから総合テスト（ユーザ確認）の間では，前段階で混入してしまったバグをなるべく多く発見し，除去する工程である．

一方で先行研究 [4, 5] では開発工程から得られるソフトウェアの規模とテスト工程で検出したバグ数を考慮した指標をソフトウェアの品質と捉え，品質を予測するモデルを作成している．先行研究の目的は予測モデルを作成する過程で品質に影響を与える変数を開発工程の中から見つけ，重点的にマネジメントすべき変数を示すことである．したがって品質の指標自体は意味を持たない．このような先行研究を受けて，重点的にマネジメントすべきフェーズの提示に留まらず，実際の開発環境にも用いることができるソフトウェアの品質や信頼性の予測ができないかと考えた．これを受けて本研究では，ソフトウェアリリース後の不具合有無をソフトウェアの信頼性とし，これを予測するモデルの作成を行うことにした．つまり本研究は，先行研究のように重点的にマネジメントすべき変数の発見を目的とするのではなく，正確なソフトウェアの信頼性予測を行ってた．

しかし，本研究の信頼性予測に用いることにした対象データは多くの無回答，つまり欠測値を含んでいた．本研究の対象データについての詳細は第２章にて説明する．したがって対象データを信頼性予測に用いるには欠測値を除外する必要がある．しかし，欠測値を除外することでデータに偏りが生じてしまったり，欠測値に伴って観測値も除外してしまうことになる．このような状況を受け，欠測を含んでいても有用な信頼性予測が可能なことを示せないかと考えた．したがって本研究の目的は，欠測値を含んでいても有用な信頼性予測が可能なことを示すことである．研究目的を達成するために２つのデータを用意した．元のデータから予測に用いるために抽出された完全データと，擬似完全データである．擬似完全データとは完全データと類似した不完全データに対して欠測値代入を行うことで擬似的に完全データとしたものである．２つのデータに対して複数の信頼性予測モデルを適用することで得られた予測精度を比較し，差がないことを示すことができれば擬似完全データからの信頼性予測は有用であると判断した．

本研究の最終的に期待する成果は，完全データと擬似完全データの両者において予測精度が高く，かつ予測精度が同じになるような対象データと信頼性予測モデルの構築である．

1.2　本論文の構成

本研究の目的はソフトウェア信頼性の予測を行う際に，不完全データからでも完全データと遜色のない予測精度を示すことである．この目的を達成するために第 2 章では元となるデータを加工することで完全データを作成し，欠測率と欠測パターンを考慮することで不完全データを作成した．第 3 章では不完全データに対して欠測値代入することで擬似的に完全データを作成できることを示し，欠測値代入法の一般論についても述べた．第 4 章では本研究で用いた欠測値代入法である EMB アルゴリズムについて説明した．第 5 章では対象データである欠測値代入された擬似完全データと完全データを作成した．第 6 章では得られた対象データの予測精度の求め方について触れた．第 7 章では予測モデルに用いた 3 つの手法について説明した．最後に，第 8 章では予測モデルから得られた予測結果を列挙，完全データと擬似完全データから得られた予測精度を比較した後に考察を述べた．

2　ソフトウェアプロジェクトデータ

本研究では元となるソフトウェアプロジェクトデータに対して条件の設定やデータ内の使用する変数の限定をすることで対象となるデータを構築した．この章では対象データの構築過程を示す．

元となるデータは IPA/SEC[*1] という独立行政法人が 2004 年から 2012 年の間，国内の約 30 のソフトウェア開発企業から収集したソフトウェアプロジェクトデータである [2]．企業に Excel ファイルのアンケートを配布し，記入を募ることで情報を収集している．このようにして収集されたデータは 3089 件のソフトウェア開発プロジェクトデータから成り，各データがソフトウェア開発情報を 611 件の変数として保持している．以上から分かる通り，元となるデータは数が非常に多い．したがって条件を設けることでデータを抽出した．そのデータの条件が以下である．

1. IPA/SEC が設定した A〜D の選択肢からなるデータの信頼性が A（データに合理性があり，完全に整合していると認められる）もしくは B（基本的に合理性があると認められるが，データの整合性に影響を及ぼす要因が幾つか存在する）である．
2. 開発プロジェクトの種別が新規開発である．新規開発とはベースとなるシステムが存在せず，新規の開発を行うもののこと．
3. 開発プロジェクトの形態が受託開発である．受託開発とは顧客が利用もしくは販売する製品の開発を請け負うことである．
4. ソフトウェア開発である．
5. 開発した情報システムの種別がアプリケーションソフトである．アプリケーションソフトとはある特定の目的を達成するためのソフトウェアのことである．応用ソフトとも呼ぶ．
6. 開発ライフサイクルモデルがウォーターフォールモデルである．ウォーターフォールモデルとはソフトウェア開発を水が上から下に落ちるように開発を順番に進めるソフトウェア開発手法である．図 1 が示しているソフトウェア開発工程もウォーターフォールモデルである．
7. 開発工程において開発 5 工程である「基本設計」,「詳細設計」,「製作」,「結合テスト」,「総合テスト（ベンダ確認）」の 5 つが必ず行われている．
8. 全体システムである．全体システムとはシステム全体のことである．したがってサブシステムの開発のようなシステムの 1 部の開発でないことを示している．
9. 主となる開発対象のプラットフォームが「WindowsNT/2000/XP 系」である．これらは OS の名前である．
10. 「業種」,「業務種類」,「主開発言語」,「業務パッケージの利用有無」,「アーキテクチャ」,「Web 技術の利用」,「DBMS の利用」のアンケート回答が欠測していない．「業種」,「業務種類」とはソフトウェアが利用される業種と業務種類の解答項目である．「主開発言語」とはソフトウェアの主な開発言語である．「アーキテクチャ」とはコンピュータのハードウェア，ソフトウェアの内部構成やそのもの自体の考え方のことである．アンケート回答の選択項目ではスタンドアロン，メインフレーム，イントラネットなどがある．「Web 技術の利用」とは，HTML のようなウェブを介して用いられる技術の利用についての項目である．「DBMS の利用」の DBMS とは Data Base Management System の略でありデータを共有するデータベースを構築するソフトウェアのことである．DBMS には様々な種類があることから「DBMS の利用」は種類について回答する項目である．

[*1] Information-Technology Promotion Agency, Japan/ Software Reliability Enhancement Center

このように条件を付け加える目的は２つある．１つ目は条件１からも分かる通りデータの信頼性の向上である．２つ目は開発プロジェクトの種別や形態などの規格を統一することで，信頼性予測の際に利用する変数が取りうる値の範囲をある程度そろえる目的もある．

次に信頼性予測に用いる変数について説明する．本研究では信頼性の定義をソフトウェアリリース後の不具合有無と決めた．これは 611 件ある変数の中から「ソフトウェアリリース後１ヵ月以内の不具合数」を用いて不具合有無を表す「0」,「1+」の２カテゴリの離散変数に変換した．この変数

y: ソフトウェアリリース後１ヵ月以内の不具合有無.

を他の変数で予測するための目的変数とした．また，この目的変数に対して説明変数を

x_1: SLOC.
x_2: 計画月数.
x_3: 平均要員数.

と定めた．SLOC とは Source Lines of Code の略であり，開発したソフトウェアの行数を表している．したがってこの指標はソフトウェアの規模と考えることができる．また，計画月数とはプロジェクトが始まる前に計画されたプロジェクトの期間であり，平均要員数とはプロジェクトに参加した各月の平均人数を表している．この３つの変数は先行研究 [5] と変数の欠測状態，変数がソフトウェア開発において一般的であるかどうかを考慮して選択した．

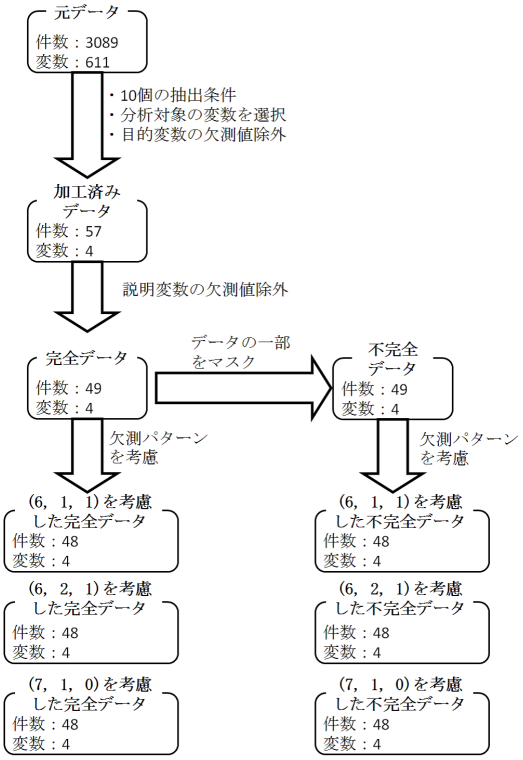

図２　完全データと不完全データの作成手順.

図 2 は不完全データと完全データの作成手順を示している．元のデータは 10 個の抽出条件の設定と 4 つの

変数選択を施した．加えて，本研究では信頼性予測が正しいかどうかを確認するため，先ほどのデータの目的変数は欠測値を含まない完全データとして取り出した．一方で説明変数は欠測値を含む不完全データとした．これが対象データのたたき台となる加工済みデータである．表 1 は加工済みデータを表したもので，表 1 内の ID とは 3089 件の元データに付記されている識別番号のことである．表 1 の通り加工済みデータの欠測値の数は左から順に (7, 2, 1) である．また，加工済みデータが持つ説明変数の欠測値を除外したものを完全データとした．この結果，データ数は 57 件から 49 件まで減少した．表 2 に完全データを示した．この完全データをマスクする（観測値を欠測値とみなす）ことで不完全データを作成した．これらの完全データと不完全データに対して欠測パターンを考慮することで 3 種類の別のデータを各々のデータから作成した．

この章ではデータを作成する過程で適用した分布変換と検定，欠測率，欠測パターンについて説明し，最後に欠測パターンを考慮した完全データと不完全データを示す．

表１　加工済みデータ.

ID	x_1	x_2	x_3	y
8	50100	4.6	2.3	1+
9	69300	7.27	1.6	0
11	56000	4.27	2.5	1+
452	71000	7.87	9.8	0
1014	370200	17	15.5	1+
1015	58500	17.93	0.79	1+
1018	2200230	30	36.4	1+
1027	12806	2.9	3.8	0
1296	N/A	16	11.9	1+
1302	273000	7.97	5.5	1+
1304	99000	24.07	0.5	1+
1307	530000	7.9	5	0
1310	N/A	N/A	1.9	0
1311	393773	N/A	N/A	1+
1312	40100	6	5	1+
1313	2800	2	16	0
1316	145500	18	1.7	0
1318	4761	4	0.9	0
1324	74160	11	1.1	1+
1325	12800	7	0.5	0
1419	280000	11	2	0
1451	N/A	4.73	2.5	0
1511	N/A	11	6	0
1526	324000	15	23	1+
1527	10585	9	2.5	0
1528	603000	6.9	7	1+
1535	10585	9	2.5	0
1542	19800	4	2	0
1629	4704	11	0.06	1+
1642	818828	15	9.5	1+

ID	x_1	x_2	x_3	y
1648	3237	5.37	2	0
1668	5600	5.67	9.73	0
1683	21700	5	3	0
1687	7800	9.5	4	0
1688	40000	4.97	7.3	0
1692	41800	8.23	2	0
1696	22200	4	3	0
1776	258000	13.97	6.4	1+
1786	11300	2.97	3.3	1+
1800	11595	3.67	2	0
1845	N/A	5.93	3	1+
1851	N/A	4	2.6	1+
2080	7903	3.87	1	1+
2098	N/A	18.97	3	1+
2341	10400	3.67	2.7	0
2346	3200	2.97	1.5	0
2538	53495	6	0.5	1+
2576	80807	5	5	0
2606	18000	5.07	1.5	1+
2789	18677	4.43	1	0
2827	43900	11.23	3.2	1+
2919	50400	7.97	10	1+
2920	92200	21	11.2	1+
2921	68400	16.97	5.9	1+
2925	362500	10.13	3.6	1+
3034	111300	11.67	5	1+
3082	65000	3.8	12	1+
欠測数	7	2	1	0

N/A　：欠測値

表 2　完全データ.

ID	x_1	x_2	x_3	y
8	50100	4.6	2.3	1+
9	69300	7.27	1.6	0
11	56000	4.27	2.5	1+
452	71000	7.87	9.8	0
1014	370200	17	15.5	1+
1015	58500	17.93	0.79	1+
1018	2200230	30	36.4	1+
1027	12806	2.9	3.8	0
1302	273000	7.97	5.5	1+
1304	99000	24.07	0.5	1+
1307	530000	7.9	5	0
1312	40100	6	5	1+
1313	2800	2	16	0
1316	145500	18	1.7	0
1318	4761	4	0.9	0
1324	74160	11	1.1	1+
1325	12800	7	0.5	0
1419	280000	11	2	0
1526	324000	15	23	1+
1527	10585	9	2.5	0
1528	603000	6.9	7	1+
1535	10585	9	2.5	0
1542	19800	4	2	0
1629	4704	11	0.06	1+
1642	818828	15	9.5	1+
1648	3237	5.37	2	0
1668	5600	5.67	9.73	0
1683	21700	5	3	0
1687	7800	9.5	4	0
1688	40000	4.97	7.3	0
1692	41800	8.23	2	0
1696	22200	4	3	0
1776	258000	13.97	6.4	1+
1786	11300	2.97	3.3	1+
1800	11595	3.67	2	0
2080	7903	3.87	1	1+
2341	10400	3.67	2.7	0
2346	3200	2.97	1.5	0
2538	53495	6	0.5	1+
2576	80807	5	5	0
2606	18000	5.07	1.5	1+
2789	18677	4.43	1	0
2827	43900	11.23	3.2	1+
2919	50400	7.97	10	1+
2920	92200	21	11.2	1+
2921	68400	16.97	5.9	1+
2925	362500	10.13	3.6	1+
3034	111300	11.67	5	1+
3082	65000	3.8	12	1+

2.1　分布変換と検定

本研究では欠測値代入を適用する変数が正規分布に従っていることを仮定している．しかし説明変数のヒストグラムである図 3, 4, 5 から分かる通り，これらの変数は正規分布に従っていない．このような事実を受けて，分布変換の一つである Box-Cox 変換 [6] を行うことで説明変数を正規分布に変換した．この節では変換過程を示す．

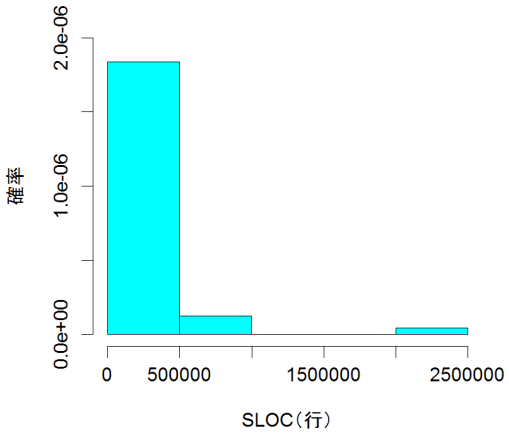

図 3　SLOC のヒストグラム．

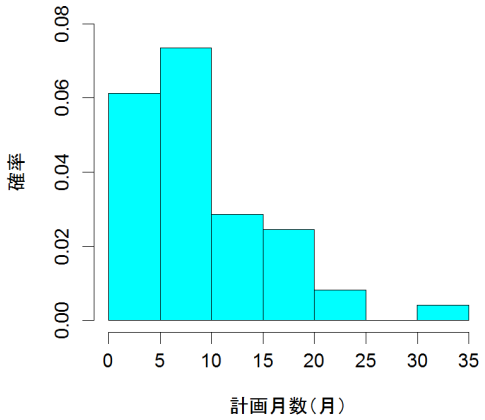

図 4　計画月数のヒストグラム．

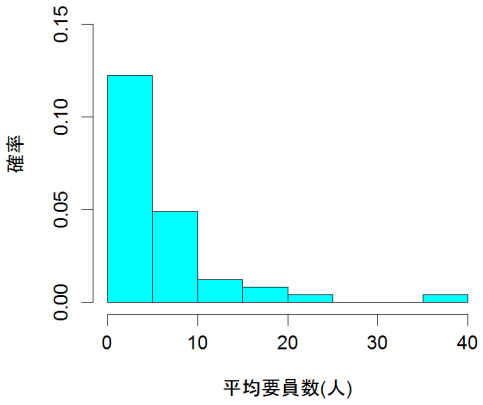

図 5　平均要員数のヒストグラム．

Box-Cox 変換とは説明変数 $x_i,\ (i=1,\ 2,\ 3)$ に

$$x_i^{(b)} = \begin{cases} \dfrac{x_i - 1}{\lambda} & (\text{if } \lambda \neq 0)\ , \\ \ln x_i & (\text{if } \lambda = 0)\ , \end{cases} \tag{2.1}$$

の式を適用させて分布変換することである．このパラメータ λ の定め方は $x^{(b)}$ の分布が $x^{(b)} \sim N(\mu, \sigma^2)$ となるように λ を最尤法によって求めた．この最尤法の計算は R 言語で行った．その結果が表 3 である．これら

表 3　最尤法により求めた各変数の λ.

変数	x_1	x_2	x_3
λ	-0.077	-0.169	0.136

の変数を用いて信頼性予測を行うのだが，λ を統一しなければ信頼性予測の際に正しい結果が得られない．したがって, 表 3 の結果を考慮しながら λ の値を変化させる（チューニングする）ことで，これらの変数が正規分布に従うような λ を実験的経験に基づいて求めた．また，正規分布に従っているかどうかは Shapiro-Wilk 検定を用いて判断した．

Shapiro-Wilk 検定とは標本 $X = (x_1, x_2, \cdots, x_n)$ が正規母集団からサンプリングされたものであるという帰無仮説を検定する検定法である [7].

帰無仮説 (H_0)：変数の分布は正規分布である．

この検定の検定統計量 W は

$$W = \frac{(a'X)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\ , \tag{2.2}$$

$$a' = (a_1, a_2, \cdots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}\ , \tag{2.3}$$

$$m' = (E(x_1) = m_1, m_2, \cdots, m_n)\ , \tag{2.4}$$

$$V = \mathrm{Cov}(x_i, x_j) = v_{ij}\ , \tag{2.5}$$

によって得られる．こうして得られた W を付録 A の Shapiro-Wilk の検定表と照合することで P 値を求めた．次に，説明変数の分布変換に用いる λ を求める過程を示す．

1. 各説明変数に式 (2.1) の変換を施した場合の尤度を最大化するような λ を最尤法により求める．
2. 得られた λ を用いて各説明変数を Box-Cox 変換する．
3. 各変数の λ を統一させるために計算過程 1 で得られた λ に基づいて新しい λ を設定する．
4. 変換後の変数を Shapiro-Wilk 検定することで分布の正規性の確認を行う．
5. もし，Box-Cox 変換後の各説明変数の分布の正規性が確認できた場合その λ を採用する．確認できない場合には計算過程 3 に戻る．

この計算過程を通して $\lambda = 0.1$ が得られた．表 4 は Box-Cox 変換と Shapiro-Wilk 検定のパラメータをまとめたものである．

表 4　Box-Cox 変換と Shapiro-Wilk 検定のパラメータのまとめ．

変数	x_1	x_2	x_3
λ_1	-0.077	-0.169	0.136
W_1	0.979	0.984	0.988
P_1	0.522	0.719	0.907
λ_2	0.1	0.1	0.1
W_2	0.957	0.975	0.987
P_2	0.074	0.384	0.852

$P < 0.05$: 有意水準 5 % の帰無仮説を棄却する．

λ_1：最尤推定量の λ.
W_1：Box-Cox 変換 ($\lambda = \lambda_1$) 後の Shapiro-Wilk 検定の検定統計量.
P_1：W_1 の P 値.
λ_2：チューニングされた λ.
W_2：Box-Cox 変換 ($\lambda = \lambda_2$) 後の Shapiro-Wilk 検定の検定統計量.
P_2：W_2 の P 値.

表 4 の P_2 が全て $P > 0.05$ であることから有意水準 5 % の帰無仮説は棄却されず，Box-Cox 変換後の説明変数は「正規分布に従わないとはいえない」と判定することができる．正確には正規分布に従っていると断定できないが，本研究では正規分布と仮定した．このようにして得られた $\lambda = 0.1$ を用いて Box-Cox 変換した完全データが表 5 である．

表 5　$\lambda = 0.1$ の Box-Cox 変換後の完全データ.

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	19.51	1.65	0.87	1+
9	20.48	2.19	0.48	0
11	19.84	1.56	0.96	1+
452	20.56	2.29	2.56	0
1014	26.04	3.28	3.15	1+
1015	19.97	3.35	-0.23	1+
1018	33.08	4.05	4.33	1+
1027	15.75	1.12	1.43	0
1302	24.96	2.31	1.86	1+
1304	21.59	3.75	-0.67	1+
1307	27.36	2.30	1.75	0
1312	18.86	1.96	1.75	1+
1313	12.12	0.72	3.20	0
1316	22.83	3.35	0.54	0
1318	13.32	1.49	-0.10	0
1324	20.69	2.71	0.10	1+
1325	15.75	2.15	-0.67	0
1419	25.05	2.71	0.72	0
1526	25.57	3.11	3.68	1+
1527	15.26	2.46	0.96	0
1528	27.85	2.13	2.15	1+
1535	15.26	2.46	0.96	0
1542	16.89	1.49	0.72	0
1629	13.29	2.71	-2.45	1+
1642	29.02	3.11	2.52	1+
1648	12.44	1.83	0.72	0
1668	13.70	1.89	2.55	0
1683	17.14	1.75	1.16	0
1687	14.50	2.52	1.49	0
1688	18.85	1.74	2.20	0
1692	18.98	2.35	0.72	0
1696	17.20	1.49	1.16	0
1776	24.77	3.02	2.04	1+
1786	15.43	1.15	1.27	1+
1800	15.49	1.39	0.72	0
2080	14.53	1.45	0.00	1+
2341	15.22	1.39	1.04	0
2346	12.41	1.15	0.41	0
2538	19.71	1.96	-0.67	1+
2576	20.96	1.75	1.75	0
2606	16.64	1.76	0.41	1+
2789	16.74	1.60	0.00	0
2827	19.12	2.74	1.23	1+
2919	19.53	2.31	2.59	1+
2920	21.37	3.56	2.73	1+
2921	20.44	3.27	1.94	1+
2925	25.97	2.61	1.37	1+
3034	21.96	2.79	1.75	1+
3082	20.29	1.43	2.82	1+

2.2 欠測率

本研究では完全データの一部をマスクすることで不完全データを作成すると決めたが，その際にいくつか考慮するべき点が生じた．一つ目は各変数においてマスクする観測値の数．二つ目はマスクする観測値の決定方法である．この節では一つ目の問題点をするために用いた欠測率について触れ，二つ目は次節にて述べる．本研究では欠測率を $^{r}/_{n}$ と定義した．これらの変数は，

r: 全データの内の欠測値を含むデータ数.
n: データの全体の数.

である．また，

r_1: 加工済みデータ内の欠測値を含むデータ数.
n_1: 加工済みデータの全体の数.

とすると $r_1 = 8,\ n_1 = 57$ であることは表 1 から既にわかっているため，$^{r}/_{n}$ は

$$\frac{r_1}{n_1} = \frac{8}{57}\ , \tag{2.6}$$

となる．本研究では完全データをマスクすることで加工済みデータの欠測率を持つ不完全データを作成したいため，式 (2.6) の結果と同様になるように完全データに対してマスクする観測値の数を求める．ここで

r_2: マスクする観測値の数.
n_2: 完全データの全体の数.

とすると $n_2 = 49$ であることは既知であることから，

$$\frac{r_1}{n_1} = \frac{r_2}{n_2}\ , \tag{2.7}$$

に r_1, n_1, n_2 を代入することで r_2 を求めることができる．その結果 $r_2 = 6.871$ となった．r_2 はデータ数を表すことから，自然数になる．したがって，$r_2 = 6.871 \simeq 7$ と定め，完全データをマスクするデータ数は 7 と決めた．

2.3 欠測パターン

欠測パターンとは欠測値を持つデータの欠測状態のことである．例えば加工済みデータ内の欠測値を持つデータの観測有無を示している表 6 から，加工済みデータの欠測パターンは (7, 2, 1) と定めた．

表 6　加工済みデータの欠測パターン．

No.	ID	x_1	x_2	x_3
1	1296	0	1	1
2	1451	0	1	1
3	1511	0	1	1
4	1845	0	1	1
5	1851	0	1	1
6	2098	0	1	1
7	1310	0	0	1
8	1311	1	0	0
欠測パターン		7	2	1

0: 欠測，1: 観測

不完全データを作成するために完全データのうち 7 つのデータをマスクすることは前節で決めたが，具体的にマスクする観測値は決めていない．この節ではマスクする観測値を欠測パターンに基づいて決める．

前節ではマスクするデータ数を 7 個にすることで加工済みデータの欠測率と完全データから作成される不完全データの欠測率をほとんど同様にすることができると考えた．したがってデータ数が 8 個である欠測パターン (7, 2, 1) から欠測を持つデータを 1 個除外することでデータ数が 7 個となる．このような手順で完全データにマスクするデータ数を決めることができた．データを除外する方法は表 6 のデータ No. 1〜 6 のうち無作為に 1 つのデータを除外する方法と，データ No. 7，データ No. 8 を除外する 3 通りある．本研究ではデータ No. 1〜 6 のうち No. 1 のデータを除外することにした．No. 1 は無作為抽出によって選択された．この方法によって表 7〜 9 のように 3 種類の欠測パターン (6, 1, 1), (6, 2, 1), (7, 1, 0) を得た．

表 7　欠測パターン (6, 1, 1).

No.	ID	x_1	x_2	x_3
1	1296	0	1	1
2	1451	0	1	1
3	1511	0	1	1
4	1845	0	1	1
5	1851	0	1	1
6	2098	0	1	1
7	1311	1	0	0
欠測パターン		6	1	1

0: 欠測，1: 観測

表 8　欠測パターン (6, 2, 1).

No.	ID	x_1	x_2	x_3
1	1451	0	1	1
2	1511	0	1	1
3	1845	0	1	1
4	1851	0	1	1
5	2098	0	1	1
6	1310	0	0	1
7	1311	1	0	0
欠測パターン		6	2	1

0: 欠測，1: 観測

表 9　欠測パターン (7, 1, 0).

No.	ID	x_1	x_2	x_3
1	1296	0	1	1
2	1451	0	1	1
3	1511	0	1	1
4	1845	0	1	1
5	1851	0	1	1
6	2098	0	1	1
7	1310	0	0	1
欠測パターン		7	1	0

0: 欠測，1: 観測

次に完全データ内のマスクする観測値の決め方について述べる．表 6 の観測値を変数に置き換えたものが表 10 である．

表 10　表 6 を変数に置き換えた表.

No.	ID	α_i	β_i	γ_i
1	1296	N/A	β_1	γ_1
2	1451	N/A	β_2	γ_2
3	1511	N/A	β_3	γ_3
4	1845	N/A	β_4	γ_4
5	1851	N/A	β_5	γ_5
6	2098	N/A	β_6	γ_6
7	1310	N/A	N/A	γ_7
8	1311	α_8	N/A	N/A
欠測パターン		7	2	1

N/A:欠測値

α_i : 欠測値を含むデータの SLOC $(i = 8)$.
β_i : 欠測値を含むデータの計画月数 $(i = 1,\ 2, \cdots,\ 6)$.
γ_i : 欠測値を含むデータの平均要員数 $(i = 1,\ 2, \cdots,\ 7)$.

各データに着目すると，少なくとも観測値が 1 つあることが分かる．表 10 の各変数と, 完全データから欠測値以外の観測値が最も近いデータを探し，該当した完全データに対して加工済みデータの欠測部分をマスクすることにした．本研究が定義する最も近いデータとは説明変数 $X_i, (i = 1,\ 2,\ 3)$ を

$X_1 = (x_{1,1},\ x_{1,2}, \cdots,\ x_{1,n})$.
$X_2 = (x_{2,1},\ x_{2,2}, \cdots,\ x_{2,n})$.
$X_3 = (x_{3,1},\ x_{3,2}, \cdots,\ x_{3,n})$.
$j,\ k,\ l = (1,\ 2, \cdots,\ n)$.

としたとき

$$d_i = \min \begin{cases} (1 - \dfrac{x_{2,k}}{\beta_i})^2 + (1 - \dfrac{x_{3,l}}{\gamma_i})^2 & (i = 1,\ 2,\ \cdots,\ 6)\ , \\ (1 - \dfrac{x_{3,l}}{\gamma_i})^2 & (i = 7)\ , \\ (1 - \dfrac{x_{1,j}}{\alpha_i})^2 & (i = 8)\ , \end{cases} \tag{2.8}$$

の d_i を求める際に採用された x を持つデータのことである．この式 (2.8) を完全データに適用させた．式 (2.8) が最小になった x を持つ加工済みデータと完全データの ID をまとめたのが表 11 である．
この表から完全データの ID と加工済みデータの ID は重複しなかったことが分かる．このようにして完全データ内のマスクする観測値を決めた．完全データを表 11 に従ってマスクすることで作成した不完全データが表 12 である．また，この不完全データは $\lambda = 0.1$ による Box-Cox 変換済みである．

表 11　欠測値を持つ加工済みデータの ID と完全データ内のマスクするデータの ID とその d_i.

No.	加工済みデータの ID	完全データの ID	d_i
1	1296	1642	0.045
2	1451	8	0.007
3	1511	3034	0.031
4	1845	1683	0.025
5	1851	11	0.006
6	2098	2827	0.171
7	1310	1800	0.003
8	1311	1014	0.004

表 12　Box-Cox 変換済み不完全データ.

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y	ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	N/A	1.65	0.87	1+	1648	12.44	1.83	0.72	0
9	20.48	2.19	0.48	0	1668	13.70	1.89	2.55	0
11	N/A	1.56	0.96	1+	1683	N/A	1.75	1.16	0
452	20.56	2.29	2.56	0	1687	14.50	2.52	1.49	0
1015	19.97	3.35	-0.23	1+	1688	18.85	1.74	2.20	0
1018	33.08	4.05	4.33	1+	1692	18.98	2.35	0.72	0
1027	15.75	1.12	1.43	0	1696	17.20	1.49	1.16	0
1302	24.96	2.31	1.86	1+	1776	24.77	3.02	2.04	1+
1304	21.59	3.75	-0.67	1+	1786	15.43	1.15	1.27	1+
1307	27.36	2.30	1.75	0	1800	N/A	N/A	0.72	0
1312	18.86	1.96	1.75	1+	2080	14.53	1.45	0.00	1+
1313	12.12	0.72	3.20	0	2341	15.22	1.39	1.04	0
1316	22.83	3.35	0.54	0	2346	12.41	1.15	0.41	0
1318	13.32	1.49	-0.10	0	2538	19.71	1.96	-0.67	1+
1324	20.69	2.71	0.10	1+	2576	20.96	1.75	1.75	0
1325	15.75	2.15	-0.67	0	2606	16.64	1.76	0.41	1+
1419	25.05	2.71	0.72	0	2789	16.74	1.60	0.00	0
1526	25.57	3.11	3.68	1+	2827	N/A	2.74	1.23	1+
1527	15.26	2.46	0.96	0	2919	19.53	2.31	2.59	1+
1528	27.85	2.13	2.15	1+	2920	21.37	3.56	2.73	1+
1535	15.26	2.46	0.96	0	2921	20.44	3.27	1.94	1+
1542	16.89	1.49	0.72	0	2925	25.97	2.61	1.37	1+
1629	13.29	2.71	-2.45	1+	3034	N/A	2.79	1.75	1+
1642	N/A	3.11	2.52	1+	3082	20.29	1.43	2.82	1+

N/A　：欠測値

2.4　完全データと不完全データ

前節では欠測率を考慮したマスクするデータの数とマスクする観測値を欠測パターンに基づいて決めた．その後，観測値をマスクすることで不完全データを作成した．この節では不完全データ（表 12）と完全データ（表 5）から欠測パターンを考慮したデータを作成する過程を示す．

まずは元となる不完全データから作成した 3 種類の不完全データについて述べる．欠測パターン (6, 1, 1), (6, 2, 1), (7, 1, 0) は加工済みデータの欠測パターンからデータを 1 つ除外したものであることから，表 12 の不完全データからデータを 1 つ除外しなければならない．したがって，不完全データから表 11 の No. 7 を除外すると欠測パターン (6, 1, 1) の不完全データが作成できる．他の欠測パターンも同様に No. 1 を除外すると欠測パターン (6, 2, 1) の不完全データ，No. 8 を除外すると欠測パターン (7, 1, 0) の不完全データが作成される．こうして得られたデータが表 14〜15 である．

表 13　欠測パターン (6, 1, 1) の Box-Cox 変換済み不完全データ．

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	N/A	1.65	0.87	1+
9	20.48	2.19	0.48	0
11	N/A	1.56	0.96	1+
452	20.56	2.29	2.56	0
1014	26.04	N/A	N/A	1+
1015	19.97	3.35	-0.23	1+
1018	33.08	4.05	4.33	1+
1027	15.75	1.12	1.43	0
1302	24.96	2.31	1.86	1+
1304	21.59	3.75	-0.67	1+
1307	27.36	2.30	1.75	0
1312	18.86	1.96	1.75	1+
1313	12.12	0.72	3.20	0
1316	22.83	3.35	0.54	0
1318	13.32	1.49	-0.10	0
1324	20.69	2.71	0.10	1+
1325	15.75	2.15	-0.67	0
1419	25.05	2.71	0.72	0
1526	25.57	3.11	3.68	1+
1527	15.26	2.46	0.96	0
1528	27.85	2.13	2.15	1+
1535	15.26	2.46	0.96	0
1542	16.89	1.49	0.72	0
1629	13.29	2.71	-2.45	1+
1642	N/A	3.11	2.52	1+
1648	12.44	1.83	0.72	0
1668	13.70	1.89	2.55	0
1683	N/A	1.75	1.16	0
1687	14.50	2.52	1.49	0
1688	18.85	1.74	2.20	0
1692	18.98	2.35	0.72	0
1696	17.20	1.49	1.16	0
1776	24.77	3.02	2.04	1+
1786	15.43	1.15	1.27	1+
2080	14.53	1.45	0.00	1+
2341	15.22	1.39	1.04	0
2346	12.41	1.15	0.41	0
2538	19.71	1.96	-0.67	1+
2576	20.96	1.75	1.75	0
2606	16.64	1.76	0.41	1+
2789	16.74	1.60	0.00	0
2827	N/A	2.74	1.23	1+
2919	19.53	2.31	2.59	1+
2920	21.37	3.56	2.73	1+
2921	20.44	3.27	1.94	1+
2925	25.97	2.61	1.37	1+
3034	N/A	2.79	1.75	1+
3082	20.29	1.43	2.82	1+

N/A　：欠測値

表 14　欠測パターン (6, 2, 1) の Box-Cox 変換済み不完全データ.

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y	ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	N/A	1.65	0.87	1+	1648	12.44	1.83	0.72	0
9	20.48	2.19	0.48	0	1668	13.70	1.89	2.55	0
11	N/A	1.56	0.96	1+	1683	N/A	1.75	1.16	0
452	20.56	2.29	2.56	0	1687	14.50	2.52	1.49	0
1014	26.04	N/A	N/A	1+	1688	18.85	1.74	2.20	0
1015	19.97	3.35	-0.23	1+	1692	18.98	2.35	0.72	0
1018	33.08	4.05	4.33	1+	1696	17.20	1.49	1.16	0
1027	15.75	1.12	1.43	0	1776	24.77	3.02	2.04	1+
1302	24.96	2.31	1.86	1+	1786	15.43	1.15	1.27	1+
1304	21.59	3.75	-0.67	1+	1800	N/A	N/A	0.72	0
1307	27.36	2.30	1.75	0	2080	14.53	1.45	0.00	1+
1312	18.86	1.96	1.75	1+	2341	15.22	1.39	1.04	0
1313	12.12	0.72	3.20	0	2346	12.41	1.15	0.41	0
1316	22.83	3.35	0.54	0	2538	19.71	1.96	-0.67	1+
1318	13.32	1.49	-0.10	0	2576	20.96	1.75	1.75	0
1324	20.69	2.71	0.10	1+	2606	16.64	1.76	0.41	1+
1325	15.75	2.15	-0.67	0	2789	16.74	1.60	0.00	0
1419	25.05	2.71	0.72	0	2827	N/A	2.74	1.23	1+
1526	25.57	3.11	3.68	1+	2919	19.53	2.31	2.59	1+
1527	15.26	2.46	0.96	0	2920	21.37	3.56	2.73	1+
1528	27.85	2.13	2.15	1+	2921	20.44	3.27	1.94	1+
1535	15.26	2.46	0.96	0	2925	25.97	2.61	1.37	1+
1542	16.89	1.49	0.72	0	3034	N/A	2.79	1.75	1+
1629	13.29	2.71	-2.45	1+	3082	20.29	1.43	2.82	1+

N/A　：欠測値

表 15　欠測パターン (7, 1, 0) の Box-Cox 変換済み不完全データ．

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y	ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	N/A	1.65	0.87	1+	1648	12.44	1.83	0.72	0
9	20.48	2.19	0.48	0	1668	13.70	1.89	2.55	0
11	N/A	1.56	0.96	1+	1683	N/A	1.75	1.16	0
452	20.56	2.29	2.56	0	1687	14.50	2.52	1.49	0
1015	19.97	3.35	-0.23	1+	1688	18.85	1.74	2.20	0
1018	33.08	4.05	4.33	1+	1692	18.98	2.35	0.72	0
1027	15.75	1.12	1.43	0	1696	17.20	1.49	1.16	0
1302	24.96	2.31	1.86	1+	1776	24.77	3.02	2.04	1+
1304	21.59	3.75	-0.67	1+	1786	15.43	1.15	1.27	1+
1307	27.36	2.30	1.75	0	1800	N/A	N/A	0.72	0
1312	18.86	1.96	1.75	1+	2080	14.53	1.45	0.00	1+
1313	12.12	0.72	3.20	0	2341	15.22	1.39	1.04	0
1316	22.83	3.35	0.54	0	2346	12.41	1.15	0.41	0
1318	13.32	1.49	-0.10	0	2538	19.71	1.96	-0.67	1+
1324	20.69	2.71	0.10	1+	2576	20.96	1.75	1.75	0
1325	15.75	2.15	-0.67	0	2606	16.64	1.76	0.41	1+
1419	25.05	2.71	0.72	0	2789	16.74	1.60	0.00	0
1526	25.57	3.11	3.68	1+	2827	N/A	2.74	1.23	1+
1527	15.26	2.46	0.96	0	2919	19.53	2.31	2.59	1+
1528	27.85	2.13	2.15	1+	2920	21.37	3.56	2.73	1+
1535	15.26	2.46	0.96	0	2921	20.44	3.27	1.94	1+
1542	16.89	1.49	0.72	0	2925	25.97	2.61	1.37	1+
1629	13.29	2.71	-2.45	1+	3034	N/A	2.79	1.75	1+
1642	N/A	3.11	2.52	1+	3082	20.29	1.43	2.82	1+

N/A　：欠測値

このようにして元となる不完全データからデータを除外することで新たな欠測パターンを持つ不完全データを 3 種類作成した．本研究は完全データと擬似完全データ（欠測値代入された不完全データ）を対象データとして信頼性予測を行い，予測結果に差が生じるかどうかを確認したいことから，3 種類の不完全データに対応する完全データも 3 種類作成した．それが表 17～18 である．

表 16　欠測パターン (6, 1, 1) に対応した Box-Cox 変換済み完全データ.

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	19.51	1.65	0.87	1+
9	20.48	2.19	0.48	0
11	19.84	1.56	0.96	1+
452	20.56	2.29	2.56	0
1014	26.04	3.28	3.15	1+
1015	19.97	3.35	-0.23	1+
1018	33.08	4.05	4.33	1+
1027	15.75	1.12	1.43	0
1302	24.96	2.31	1.86	1+
1304	21.59	3.75	-0.67	1+
1307	27.36	2.30	1.75	0
1312	18.86	1.96	1.75	1+
1313	12.12	0.72	3.20	0
1316	22.83	3.35	0.54	0
1318	13.32	1.49	-0.10	0
1324	20.69	2.71	0.10	1+
1325	15.75	2.15	-0.67	0
1419	25.05	2.71	0.72	0
1526	25.57	3.11	3.68	1+
1527	15.26	2.46	0.96	0
1528	27.85	2.13	2.15	1+
1535	15.26	2.46	0.96	0
1542	16.89	1.49	0.72	0
1629	13.29	2.71	-2.45	1+
1642	29.02	3.11	2.52	1+
1648	12.44	1.83	0.72	0
1668	13.70	1.89	2.55	0
1683	17.14	1.75	1.16	0
1687	14.50	2.52	1.49	0
1688	18.85	1.74	2.20	0
1692	18.98	2.35	0.72	0
1696	17.20	1.49	1.16	0
1776	24.77	3.02	2.04	1+
1786	15.43	1.15	1.27	1+
2080	14.53	1.45	0.00	1+
2341	15.22	1.39	1.04	0
2346	12.41	1.15	0.41	0
2538	19.71	1.96	-0.67	1+
2576	20.96	1.75	1.75	0
2606	16.64	1.76	0.41	1+
2789	16.74	1.60	0.00	0
2827	19.12	2.74	1.23	1+
2919	19.53	2.31	2.59	1+
2920	21.37	3.56	2.73	1+
2921	20.44	3.27	1.94	1+
2925	25.97	2.61	1.37	1+
3034	21.96	2.79	1.75	1+
3082	20.29	1.43	2.82	1+

表 17　欠測パターン (6, 2, 1) に対応した Box-Cox 変換済み完全データ.

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y		ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	19.51	1.65	0.87	1+		1648	12.44	1.83	0.72	0
9	20.48	2.19	0.48	0		1668	13.70	1.89	2.55	0
11	19.84	1.56	0.96	1+		1683	17.14	1.75	1.16	0
452	20.56	2.29	2.56	0		1687	14.50	2.52	1.49	0
1014	26.04	3.28	3.15	1+		1688	18.85	1.74	2.20	0
1015	19.97	3.35	-0.23	1+		1692	18.98	2.35	0.72	0
1018	33.08	4.05	4.33	1+		1696	17.20	1.49	1.16	0
1027	15.75	1.12	1.43	0		1776	24.77	3.02	2.04	1+
1302	24.96	2.31	1.86	1+		1786	15.43	1.15	1.27	1+
1304	21.59	3.75	-0.67	1+		1800	15.49	1.39	0.72	0
1307	27.36	2.30	1.75	0		2080	14.53	1.45	0.00	1+
1312	18.86	1.96	1.75	1+		2341	15.22	1.39	1.04	0
1313	12.12	0.72	3.20	0		2346	12.41	1.15	0.41	0
1316	22.83	3.35	0.54	0		2538	19.71	1.96	-0.67	1+
1318	13.32	1.49	-0.10	0		2576	20.96	1.75	1.75	0
1324	20.69	2.71	0.10	1+		2606	16.64	1.76	0.41	1+
1325	15.75	2.15	-0.67	0		2789	16.74	1.60	0.00	0
1419	25.05	2.71	0.72	0		2827	19.12	2.74	1.23	1+
1526	25.57	3.11	3.68	1+		2919	19.53	2.31	2.59	1+
1527	15.26	2.46	0.96	0		2920	21.37	3.56	2.73	1+
1528	27.85	2.13	2.15	1+		2921	20.44	3.27	1.94	1+
1535	15.26	2.46	0.96	0		2925	25.97	2.61	1.37	1+
1542	16.89	1.49	0.72	0		3034	21.96	2.79	1.75	1+
1629	13.29	2.71	-2.45	1+		3082	20.29	1.43	2.82	1+

表 18　欠測パターン (7, 1, 0) に対応した Box-Cox 変換済み完全データ．

ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y	ID	$x_1^{(b)}$	$x_2^{(b)}$	$x_3^{(b)}$	y
8	19.51	1.65	0.87	1+	1648	12.44	1.83	0.72	0
9	20.48	2.19	0.48	0	1668	13.70	1.89	2.55	0
11	19.84	1.56	0.96	1+	1683	17.14	1.75	1.16	0
452	20.56	2.29	2.56	0	1687	14.50	2.52	1.49	0
1015	19.97	3.35	-0.23	1+	1688	18.85	1.74	2.20	0
1018	33.08	4.05	4.33	1+	1692	18.98	2.35	0.72	0
1027	15.75	1.12	1.43	0	1696	17.20	1.49	1.16	0
1302	24.96	2.31	1.86	1+	1776	24.77	3.02	2.04	1+
1304	21.59	3.75	-0.67	1+	1786	15.43	1.15	1.27	1+
1307	27.36	2.30	1.75	0	1800	15.49	1.39	0.72	0
1312	18.86	1.96	1.75	1+	2080	14.53	1.45	0.00	1+
1313	12.12	0.72	3.20	0	2341	15.22	1.39	1.04	0
1316	22.83	3.35	0.54	0	2346	12.41	1.15	0.41	0
1318	13.32	1.49	-0.10	0	2538	19.71	1.96	-0.67	1+
1324	20.69	2.71	0.10	1+	2576	20.96	1.75	1.75	0
1325	15.75	2.15	-0.67	0	2606	16.64	1.76	0.41	1+
1419	25.05	2.71	0.72	0	2789	16.74	1.60	0.00	0
1526	25.57	3.11	3.68	1+	2827	19.12	2.74	1.23	1+
1527	15.26	2.46	0.96	0	2919	19.53	2.31	2.59	1+
1528	27.85	2.13	2.15	1+	2920	21.37	3.56	2.73	1+
1535	15.26	2.46	0.96	0	2921	20.44	3.27	1.94	1+
1542	16.89	1.49	0.72	0	2925	25.97	2.61	1.37	1+
1629	13.29	2.71	-2.45	1+	3034	21.96	2.79	1.75	1+
1642	29.02	3.11	2.52	1+	3082	20.29	1.43	2.82	1+

欠測パターンに対応した完全データの作成方法は，不完全データと同様である．表 11 の No. 7 を除外すると欠測パターン (6, 1, 1) に対応する完全データが作成できる．他の欠測パターンも同様に No. 1 を除外すると欠測パターン (6, 2, 1) に対応する完全データ, No. 8 を除外すると欠測パターン (7, 1, 0) に対応する完全データが作成される．作成した完全データは対象データであるが欠測パターンごとに変数を定義しておらず，不完全データに関しても次の章の欠測代入を経て対象データとなる．したがって対象データを確認したい場合は第 5.2 節に示すとおり，付録 B, C の完全データと擬似完全データを確認してほしい．

3　欠測値代入

欠測値代入とはデータの欠測部分に値を代入することである．欠測値に値を代入することを補定するともいう．欠測が起きてしまう理由はいくつか存在する．次に簡単な例を列挙する．

1. アンケート調査における無回答
2. 長期間の追跡に伴って生じるデータの脱落
3. 生存時間解析における打ち切り
4. 計測装置の測定限界を超えるデータ

このうち本研究で取り扱っている欠測値は「アンケート調査における無回答」によって生じている．上記にあげた以外にもデータの欠測が生じてしまう原因は多岐にわたっている．したがって調査や実験において計画された通りに完全なデータを観測できることはほとんどないと考えられる．次に欠測値代入の必要性について説明する．

表 19 は計測を目的とした何らかの変数 A, B, C の観測状況を示したものである．表 19 のデータを用いて統計解析する際に，データ No. 3, 7, 8 のように統計解析を行うデータ内に 1 つでも欠測値が生じるとそのデータ全てを除外しなければならない．ちなみに，このようなデータの除去方法を「リストワイズ法」と呼ぶ．このデータ除去が原因で捨てられてしまう観測値を表 19 内では括弧をつけて表した．欠測値を代入することができればリストワイズ法は適用しないため，括弧を持つ観測値も統計解析に組み込むことができる．したがって欠測値代入を行うことで統計解析の精度をあげることができる．

表 19　3 つの変数の観測例.

No.	A	B	C
1	1	1	1
2	1	1	1
3	(1)	0	(1)
4	1	1	1
5	1	1	1
6	1	1	1
7	0	(1)	(1)
8	(1)	0	(1)

0:欠測，1:観測

また，全米研究評議会[*2]から欠測値の処理に関するレポート [8] が出版され，日本の製薬協でもこれを基にした臨床試験の欠測値の取り扱いについての報告書がまとめられている [9]．これらのレポートは薬品の統計解析に関する欠測値の取り扱いではあるが，このような動向は他分野にも波及していくと考えられる．以上の事柄から欠測値処理の必要性は年々高まっていると判断できる．

この章では欠測値のメカニズム，単一代入法，多重代入法について説明する．

[*2] National Research Council

3.1 欠測メカニズム

この章のはじめに説明したとおり欠測値が生じる原因は様々である．しかし，欠測原因は 3 つの欠測メカニズムに分類することができる [10]．この節では 3 種類の欠測メカニズムと統計解析の際の注意点を説明する．

Missimg Completely at Random (MCAR)

欠測メカニズム MCAR のもとでは計測されている変数 X で生じる欠測は完全にランダムである．つまり，X の欠測は他の計測されている変数の値に依らず，加えて X の値にも依らない．MAR に従う変数を用いた統計解析では観測値を全ての計測データ（観測値と欠測値）のランダムサンプルとみなすことができる．したがって，欠測値が除外された完全データでの統計解析でもデータに偏りは生じないと考えられる．

Missimg at Random (MAR)

欠測メカニズム MCAR のもとでは計測されている変数 X で生じる欠測は他の計測されている変数 Y の値に影響されている．また，この変数 X は計測されていない要因には依存していないと考える．つまり，X の欠測メカニズムは計測されている変数 Y によって説明できる．したがって MAR はランダムな欠測ではないため，計測されたデータを考慮して欠測値処理を行うことで欠測値によって生じるバイアスを補正することができる．また，欠測データの統計解析の多くは MAR を想定しているため欠測値代入の理論が充実している．

Not Missimg at Random (NMAR)

欠測メカニズム NMAR のもとでは計測されている変数 X で生じる欠測は他の計測されている変数 Y だけでなく計測されていない変数 Z にも依存して起こる．つまり，X の欠測メカニズムは計測された変数 Y だけでは説明できない．NMAR は計測していない変数によって欠測が規定されてしまうため欠測処理が難しい．NMAR を仮定した統計解析ではデータに偏りが生じているため，パラメータの推定には選択モデルやパターン混合モデルなどの強い仮定をもつモデルを適用する必要がある．そのため，この欠測パターンでは統計解析におけるデータの前提条件をクリアすることが難しい．

本研究のデータ x_1, x_2, x_3, y は MAR を仮定し，欠測値処理として欠測値代入を行った．

3.2　単一代入法

欠測値代入は単一代入法と多重代入法の 2 つに分類することができる．単一代入法は 1 つの不完全データの欠測値に 1 つの値を代入することで，擬似完全データを 1 つ作成する方法である．単一代入法の中から主なものを 3 つ紹介する．説明するにあたり文献 [11, 12, 13] を参考にした．

平均値代入法

計測された変数 X のうち観測値の平均値を欠測値に代入する方法である．古典的な方法ではあるが平均値代入法は大きな欠点を 2 つ持っている． 1 つは標本平均の算出に貢献していない点である．欠測値に平均値を代入しても標本平均は変わらないため平均値を伴う統計解析では欠測値代入による結果の変化が期待できない． 2 つ目は標準偏差が過小推定されてしまう点である．図 6 は平均値 100，標準偏差が 10 の正規乱数を 1000 個を作成し，そのうち 100 個の観測値を欠測メカニズム MCAR のもとで欠測とみなし，平均値代入した正規乱数のヒストグラムである．図 6 から分かるとおり，横軸の 100 あたりで大きな偏りがあることが視覚的に分かる．また，図 6 の標準偏差は 9.683 となった．以上のことから平均値代入は標準偏差を過小推定してしまうことが分かる．

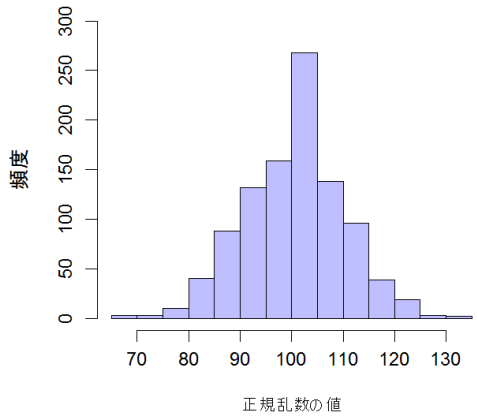

図 6　平均値代入された正規乱数のヒストグラム．

回帰代入法

データに回帰モデルをあてはめることで予測値を求め，欠測値に予測値を代入する方法である．回帰モデルの作成方法について簡単に説明する．ここでは目的変数 y を説明変数 x で予測する単回帰モデルを考える．この x は欠測値を含む不完全データとした．まず，x の観測値を持つデータのみで

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \ , \tag{3.1}$$

の α と β を最小 2 乗法により推定し，得られた推定値 $\hat{\alpha}, \hat{\beta}$ と x が欠測値であるデータの y を式 (3.1) に代入することで x の予測値，つまりは欠測値に代入する値を求めることができる．

今回は単回帰モデルで説明したが同様に重回帰モデルでも適用できる．したがって平均値代入法とは違い 1 変数のみでなく複数の変数を考慮して予測値を求めることができる．しかし，x において欠測値が複数存在し，かつ x の欠測値に対応する y の値が等しい場合，予測値も等しくなる．これは複数の欠測値に同じ予測値を代入することを意味している．したがって平均値代入法と同様に変数のばらつきを過小推定してしまっている．また完全データと不完全データに対して回帰モデルをあてはめることで 2 種類の α と β とを得ることができる．したがって完全データから求めたパラメータを真値と考えると不完全データから求めたパラメータは同一でないことが確認されている [11]．このような現象を推定不確実性と呼ぶ．回帰代入法を適用する場合，推定不確実性に関する議論が必須である．

確率的回帰代入法

回帰代入法の欠点である「変数のばらつきの過小推定」を補うために考案されたのが確率的回帰代入法である [10]．回帰代入法で求めた残差に対して推定されたランダムな誤差項を加えることで予測値を求めている．予測値の求め方について説明する．式 (3.1) を用いて残差 u を

$$\hat{u} = y - \hat{\alpha} - \hat{\beta}x \ , \tag{3.2}$$

のように求めた後に，得られた残差を用いて標準偏差 σ_u を算出する．これらの変数と標準正規乱数 z を用いて

$$\hat{y} = \hat{\alpha} + \hat{\beta}x + \sigma_u z \ , \tag{3.3}$$

とすることで不確実性を持たせることができた．確率的回帰代入法は式 (3.3) を回帰代入法の式 (3.1) のように用いて予測値を求める方法である．確率的回帰代入は変数のばらつきを考慮することはできるがモデルは 1 つであることから回帰代入と同様に推定不確実性が生じてしまう．

標準正規乱数によって回帰代入法の欠点を補うことができたが，一方で別の欠点も発生した．モデル内に標準正規乱数を組み込んでいるため解析の度に予測結果が異なってしまい，結果の再現性がない．また，ばらつきが発生しているため予測値の標準偏差を正しく求めることができない．しかしながら，回帰代入法と比べるとより自然な欠測値代入法と考えられる．

3.3　多重代入法

多重代入法は 1 つの不完全データの欠測値に異なる M 個の値を代入することで，M 個の擬似完全データの副標本を作成する．その副標本に対して目的の分析を行うことで M 個の分析結果を得ることができる．その後，M 個の分析結果を統合することにより 1 つの統合された分析結果を示すことができる [14]．$M = 5$ の多重代入法を示したものが図 7 である．

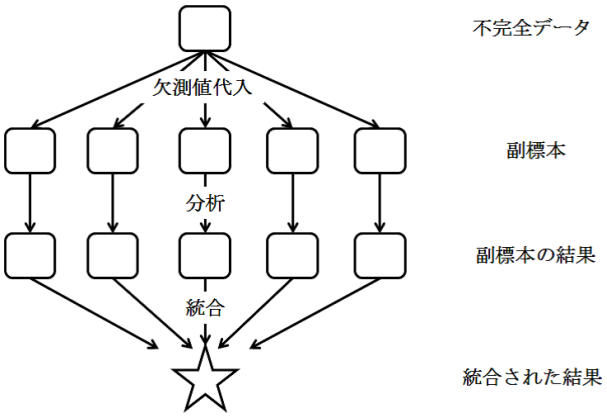

図 7　多重代入法の概要．

以上から，多重代入法を適用するにあたり重要な点を列挙する．

M（欠測値代入の数）の決定方法．
欠測値を考慮したデータのパラメータ μ, σ^2 の決定方法．
副標本の分析結果の統合方法．

M と μ, σ^2 の取り決め方法は第 4 章にて説明する．また副標本の分析結果の統合方法に関しては第 6 章で触れる．この節では代表的な多重代入法を簡単に説明する．

マルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo)

マルコフ連鎖モンテカルロ法 (MCMC) は 1 番最初に考案された多重代入法である [14]．この方法は，取り決めた確率分布に基づいたシミュレーション値を発生させる方法である．MCMC の計算アルゴリズムは多く考案されており，用途に応じて計算アルゴリズムを変更することが通常である．欠測値代入で用いるアルゴリズムはデータ拡大法 (Data Augmentation) である．データ拡大法では欠測値に適当な値の初期値 σ_0 を代入することで擬似的にデータを拡大し，擬似完全データを作成する．その後，MCMC の考え方に基づいた繰り返し手法を用いて推定値 σ を改善していく方法である．この方法は 2 ステップで構成されている．観測値を持つデータを Y_o, 欠測値を持つデータを Y_m としたとき，

I-step: $P(Y_m|Y_o, \sigma_t)$ に基づいて $Y_m^{(t+1)}$ を作成する．
P-step: $P(\sigma|Y_o, Y_m^{(t+1)})$ に基づいて σ_{t+1} を作成する．

である．I-step と P-step を推定値 σ が収束するまで繰り返す．また，t は繰り返し回数を意味している．こうして得られた収束値 σ_t が欠測値に代入する値となる．

完全条件付き指定 (Fully Conditional Specification)

完全条件付き指定（FCS）は 2 変量のデータに適用することはできない．欠測値を持つ各々の変数に対して欠測値代入モデルを作成し，各変数に対して欠測値代入を繰り返している [15]．したがって欠測値を持つ多変量データにのみ適用することができる．またこれらの変数の中から 1 つを取り出したとき, 残りの他の変数による条件付き分布でパラメトリックに特定できることが前提とされている．欠測値代入モデルは複数存在するがここではその中の 1 つを紹介する．欠測値を持つ変数を $x_1, x_2, \cdots, x_p$ とする．

1. それぞれの変数の欠測値を，同一の変数で観測された他の値に置き換えることで擬似完全データを作成する．このように置き換えた値が FCS の初期値である．
2. 得られた擬似完全データに対して $x_i (i = 1, 2, \cdots, p)$ のうち，元々欠測値だった値を除外する．
3. 確率的回帰代入法を用いて予測値を求める．
4. 得られた予測値を除外していた x_i の欠測値に代入することで擬似完全データを更新する．
5. $i = i + 1$ とし，2. に戻る．

この計算手順を全ての変数で行ったあともう 1 度初めの変数に戻ることで計算を繰り返す．この繰り返し回数を j とすると，この j の回数が欠測値代入済みの擬似完全データの数に相当する．しかし，はじめの 10〜20 回は代入値が収束していないことから，切り捨てる場合が多い．これを Burn-In という．

EMB(Expectation-Maximization with Bootstrapping) アルゴリズム

EM アルゴリズムにノンパラメトリックブートストラップ法を応用したものである [16]．本研究ではこの欠測値代入法を採用した．第 4 章にて詳しく説明する．

4　EMB アルゴリズム

EMB アルゴリズムとは EM アルゴリズムとノンパラメトリックブートストラップを応用した多重代入法である．図 8 は EMB アルゴリズムの概要を示している．用意した不完全データに対してノンパラメトリックブートストラップを行うことで欠測値を含む副標本を作成する．その後，各副標本に対して EM アルゴリズムを適用することで欠測値代入を行っている．

最適な M の数については現在も多く議論がなされている課題である．1 部の文献では欠測率が極端に高くない限り M は 5〜10 程度でよいとされている [17]. しかし，その根拠は示されていないことが多い．一方で現在のコンピュータの計算速度は飛躍的に向上しており，多重代入法の計算も容易に行えることから副標本の数，つまりは欠測値代入回数 M は数百以上であっても良いという意見もある [13]．以上のような意見を受け，本研究では副標本の数 M は 5 つと定めた．このように取り決めた理由は，図 8 内の結果の統合方法が M が 100 以上になってしまうと計算時間が非常にかかってしまうことから，前者の意見を参考に M を 5 と決めた．

このような EMB アルゴリズムの計算は R 言語の ‘Amelia ’ というパッケージを用いて行った [18]．Amelia は 2 つの前提を設けている．1 つ目は完全データは多変量正規分布であること．2 つ目は，欠測のメカニズムが MAR または MAR であることを想定している．第 2.1 節で対象データは正規分布に変換されていることから 1 つ目の前提はクリアしている．加えて，対象データは MAR を想定していることから 2 つ目の前提もクリアしている．この章では EMB アルゴリズムで用いられているノンパラメトリックブートストラップと EM アルゴリズムについて説明する．

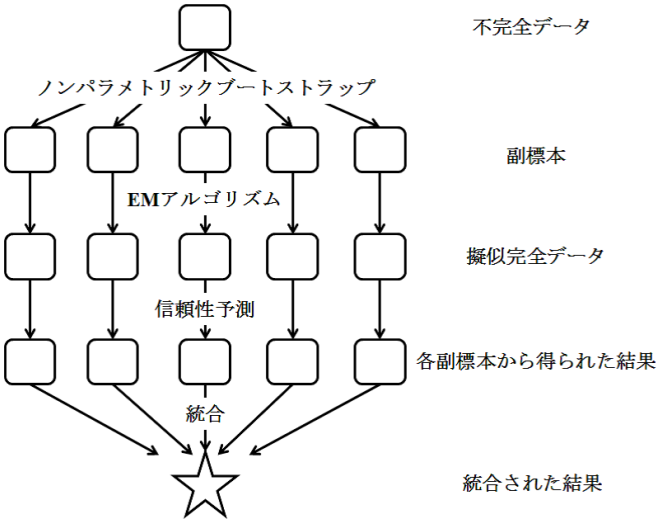

図 8　EMB アルゴリズムの概要.

4.1　ノンパラメトリックブートストラップ

ノンパラメトリックブートストラップ法では，標本サイズ n の観測された標本データから，標本サイズ n の副標本 (subsample) の無作為な復元抽出（重複を許す抽出）を行っている．本研究ではこの方法で副標本を作成している．

4.2 EM アルゴリズム

本研究で適用した EM アルゴリズムについて考える．EM アルゴリズムとは不完全な状態で観測されたデータについて最尤法に基づいた推測を行うための方法であり，E ステップと M ステップに計算が分かれている [16]．E ステップとは欠測について観測値が与えれたという条件のもとでその条件付期待値を求め，欠測値に擬似的な観測値として代入する操作のことである．M ステップとは欠測値を条件付期待値で置き換えて得られた擬似的な完全データに基づいて完全な観測に関する尤度を最大化する操作のことである．この EM アルゴリズムを本研究では欠測値代入法として用いた．この節では本研究で適用した 1 変量正規分布を仮定した EM アルゴリズムの計算手順を説明する．EM アルゴリズムの計算は文献 [19, 20] を参照した．

独立に正規分布 $N(\mu, \sigma^2)$ に従う確率変数 X_i $(i = 1, 2, \cdots, n)$ の観測を計画したと考える．このときにランダムに欠測が発生し，n 個の観測予定データのうち m 個のみが観測されたとする．標記の簡便化のため，前半の m 個のデータは観測され，後半の m' $(= n - m)$ 個については欠測してしまった状態だとする．したがって完全に観測の行われた確率変数ベクトルを $\boldsymbol{X} = (X_1, \cdots, X_m, X_{m+1}, \cdots, X_n)^T$ とすると実際に観測の行われた標本に対する確率変数ベクトル $\boldsymbol{Y}$ は，$\boldsymbol{Y} = (Y_1, \cdots, Y_m)^T$ とし，その実現値を $\boldsymbol{y} = (y_1, \cdots, y_m)^T$ と表すことにする．また，欠測データに関する確率変数は $\boldsymbol{Z} = (X_{m+1}, \cdots, X_n)^T = (Z_1, \cdots, Z_{m'})^T$ と表すことにした．このとき，適当なパラメータの推定値を初期値 $\boldsymbol{\theta^{(0)}}$ とした．観測の行われた m 標本の観測値 y に基づく対数尤度 $l(\boldsymbol{\theta},\ \boldsymbol{y})$ について

$$l(\boldsymbol{\theta},\ \boldsymbol{y}) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \mu)^2\ , \tag{4.1}$$

となる．ただし，$\boldsymbol{\theta} = (\mu, \sigma^2)^T$ である．仮に欠測値が観測されたと想定して得られた完全データ x に関する対数尤度 $l^C(\boldsymbol{\theta}, \boldsymbol{x})$ は同様に

$$\begin{aligned} l^C(\boldsymbol{\theta},\ \boldsymbol{x}) &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}(y_i - \mu)^2 + \sum_{i=1}^{m'}(z_j - \mu)^2\right\}\ , \end{aligned} \tag{4.2}$$

となる．式 (4.2) 内の $z = (z_1, \cdots, z_{m'})$ は欠測 Z に関する実際には存在しない仮想した実現値である．ここで，EM アルゴリズムの E ステップの k 段階目の計算でパラメータ推定値 $\boldsymbol{\theta^{(k)}}$ が得られたとする．このとき $(k+1)$ 段階目の E ステップでは，観測値である $\boldsymbol{Y} = \boldsymbol{y}$ のもとでの $\boldsymbol{\theta} = \boldsymbol{\theta^{(k)}}$ のときの Z の条件分布に基づく期待値の計算が必要になる．このデータは独立標本を考えており欠測もランダムであることから，確率変数 Z_i の観測 $\boldsymbol{Y} = \boldsymbol{y}$ のもとでの $l^C(\boldsymbol{\theta}, \boldsymbol{X})$ の条件付期待値 $Q(\boldsymbol{\theta}, \boldsymbol{\theta^{(k)}})$ は $\boldsymbol{\theta^{(k)}} = (\mu^{(k)}, (\sigma^2)^{(k)})$ を用いて

$$\begin{aligned} Q(\boldsymbol{\theta},\ \boldsymbol{\theta}^{(k)}) &= E_{\boldsymbol{\theta}^{(k)}}[l^C(\boldsymbol{\theta}, \boldsymbol{X})|\boldsymbol{Y} = \boldsymbol{y}] \\ &= E_{\boldsymbol{\theta}^{(k)}}\left[-\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}(Y_i - \mu)^2 + \sum_{j=1}^{m'}(Z_j - \mu)^2\right\}\middle| Y = y\right] \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}(y_i - \mu)^2 + \sum_{j=1}^{m'}E_{\boldsymbol{\theta^{(k)}}}[(Z_j - \mu)^2|\boldsymbol{Y} = \boldsymbol{y}]\right\} \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}(y_i - \mu)^2 + m'(\mu^{(k)} - \mu)^2 + m'(\sigma^2)^{(k)}\right\}, \end{aligned} \tag{4.3}$$

と表すことができる．ここまでの $Q(\boldsymbol{\theta},\ \boldsymbol{\theta^{(k)}})$ を求める操作を E ステップと呼ぶ．次の M ステップでは $Q(\boldsymbol{\theta}, \boldsymbol{\theta^{(k)}})$ を最大化するような θ を求めることになる．この θ は

$$
\begin{aligned}
0 = \frac{\partial}{\partial\theta}Q(\theta,\theta^{(k)}) &= \begin{bmatrix} \dfrac{\partial}{\partial\mu}Q(\theta,\theta^{(k)}) \\ \dfrac{\partial}{\partial\sigma^2}Q(\theta,\theta^{(k)}) \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{1}{\sigma^2}\left\{\displaystyle\sum_{i=1}^{m}(y_i-\mu)+m'(\mu^{(k)}-\mu)\right\} \\ -\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\left\{\displaystyle\sum_{i=1}^{m}(y_i-\mu)^2+m'(\mu^{(k)}-\mu)^2+m'(\sigma^2)^{(k)}\right\} \end{bmatrix},
\end{aligned} \tag{4.4}
$$

を解くことにより，$(k+1)$ 段階目のパラメータ $\theta^{(k+1)}$ の推定値を

$$
\sigma^{(k+1)} = \begin{bmatrix} \mu^{(k+1)} \\ (\sigma^2)^{(k+1)} \end{bmatrix} = \begin{bmatrix} \dfrac{1}{n}\left\{\displaystyle\sum_{i=1}^{m}y_i+m'\mu^{(k)}\right\} \\ \dfrac{1}{n}\left\{\displaystyle\sum_{i=1}^{m}(y_i-\mu^{(k+1)})^2+m'(\mu^{(k)}-\mu^{(k+1)})^2+m'(\sigma^2)^{(k)}\right\} \end{bmatrix}, \tag{4.5}
$$

によって求めることができる．

また，EM アルゴリズムで求めたパラメータのベクトル $\sigma^{(k)} = 1, 2, \cdots, k$ が収束するまで E ステップと M ステップを繰り返す．収束した場合には十分大きな k で

$$
\theta^{(k+1)} = \theta^{(k)}\ , \tag{4.6}
$$

が成立することが分かっている [19]．したがって式 (4.6) の状態のパラメータ σ を $\sigma^{(*)}$ として，パラメータの更新式である式 (4.5) に代入すると，

$$
\mu^{(*)} = \frac{1}{n}\left\{\sum_{i=1}^{m}y_i+m'\mu^{(*)}\right\}\ , \tag{4.7}
$$

$$
(\sigma^2)^{(*)} = \frac{1}{n}\left\{\sum_{i=1}^{m}(y_i-\mu^{(*)})^2+m'(\sigma^2)^{(*)}\right\}\ , \tag{4.8}
$$

が成立する．したがって $\mu^{(*)},\ (\sigma^2)^{(*)}$ は

$$
\mu^{(*)} = \frac{1}{m}\sum_{i=1}^{m}y_i\ , \tag{4.9}
$$

$$
(\sigma^2)^{(*)} = \frac{1}{m}\sum_{i=1}^{m}(y_i-\mu^{(*)})^2\ , \tag{4.10}
$$

と表現することができる．以上の手順で得られた最尤推定量 $\theta^{(*)}$ の $\mu^{(*)}$ が欠測値に代入する値である．‘Amelia’ では EM アルゴリズムにおいてもノンパラメトリックブートストラップ法を用いることで EM アルゴリズムの対象データが分析ごとに変動する．したがって欠測値に代入される値も分析ごとに異なる．

5　擬似完全データの作成

第 4 章では欠測値代入法である EMB アルゴリズムの一般論を説明した．本研究では EMB アルゴリズムを第 2.4 節で紹介した不完全データ（表 14〜表 15）に実際に適用させることで信頼性予測の対象データである擬似完全データを作成した．図 9 が示す通り，擬似完全データは 3 種類の不完全データに欠測値代入することで作成されていることから，擬似完全データの作成過程は図 2 の続きであることを示している．この章では EMB アルゴリズムを用いた欠測値代入の過程とそれによって得られた擬似完全データを紹介する．

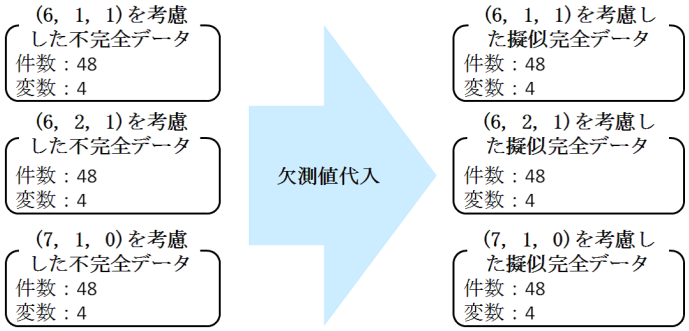

図 9　擬似完全データの作成．

5.1　欠測値代入過程

EMB アルゴリズムを用いた欠測値代入の過程を説明する．

手順 1

欠測パターンを考慮した不完全データ（表 13 〜表 15 のうちの一つ）を用意する．

手順 2

不完全データに対してノンパラメトリックブートストラップを用いて欠測値を含む副標本を 5 つ作成する．

手順 3

各副標本に対して EM アルゴリズムを適用し，欠測値代入を行うことで擬似完全データが 5 つ作成される．これら 5 つの擬似完全データは I，II，III，IV，V で表す．

この計算手順を 3 種類の不完全データに対して行うため，15 の擬似完全データが作成される．次節にて擬似完全データを紹介する．

5.2 対象データ

第 2.4 節では 3 種類の完全データを紹介した（表 16〜表 18）．また，第 5.1 節では 15 個の擬似完全データを作成した．この節ではこれらの完全データと擬似完全データ内の変数を定義する．また，完全データである表 16〜表 18 は不完全データ作成のための利用に留まっていたため，対象データとして変数の定義がなされていない．したがって，この節内で改めて対象データの変数として定義し直す．

i: 欠測パターン, $\{i = (6,\ 1,\ 1), (6,\ 2,\ 1), (7,\ 1,\ 0)\}$.
x_1^i: 欠測パターン i に対応した完全データの $x_1^{(b)}$.
x_2^i: 欠測パターン i に対応した完全データの $x_2^{(b)}$.
x_3^i: 欠測パターン i に対応した完全データの $x_3^{(b)}$.
y^i: 欠測パターン i に対応した完全データの y.

j: 欠測パターン毎の擬似完全データの識別番号 ($j =$ I，II，III，IV，V).
$x_{(1,j)}^i$: 欠測パターン i に対応した擬似完全データ j の $x_1^{(b)}$.
$x_{(2,j)}^i$: 欠測パターン i に対応した擬似完全データ j の $x_2^{(b)}$.
$x_{(3,j)}^i$: 欠測パターン i に対応した擬似完全データ j の $x_3^{(b)}$.
y_j^i: 欠測パターン i に対応した擬似完全データ j の $y^{(b)}$.

これらの変数を用いて対象データである完全データと擬似完全データを表現するが，3 種の完全データと 15 個の擬似完全データを節内で説明するには量が多いため付録 B, C として巻末に記載した．

6　信頼性予測

本研究では y が 0 の場合には信頼性が高いソフトウェアだと考え，1+ のときに信頼性が低いと考える．したがって信頼性の予測とは y を予測することである．予測する際にモデルを利用するが，予測モデルを構築するための学習データと y の予測値と真値を比較するためのテストデータが必要となる．この 2 つのデータは対象データから取り出される．学習データとテストデータの決定方法は 2 種類あり，第 6.1 節で詳しく説明する．表 20 は，ある欠測パターンに対応した完全データから得た予測値，表 21 は，ある欠測パターンの擬似完全データから求めた予測値をまとめた表 21 から分かる通り擬似完全データは欠測パターンごとに I〜V まであることから各予測値の多数決をとることで結果を統合し，$\hat{y}$ を求めた．

学習データから作成された予測モデルを用いてテストデータの y を予測し，真値と予測値を比較することで正答率を求めることができる．表 20 と表 21 の表記に従うと，y と $\hat{y}$ を比較することである．この正答率をモデルの予測精度とする．欠測パターンごとの完全データから得られた予測精度と擬似完全データから同様に得られた予測値を比較することで予測精度の差を求める．仮に表 20 と表 21 において $n = 10$ としたときに表 20 の予測精度が 6/10，表 21 の予測精度が同様に 6/10 の場合は予測精度に差がないと判断する．差がないことを確認できれば擬似完全データからでも完全データと遜色のない信頼性予測が可能だとみなした．

信頼性予測における最終目標は，完全データと擬似完全データの予測精度に差がなく且つ信頼性予測に足りうる予測精度を示すことができる予測モデルの構築である．この章では学習データとテストデータの決定と信頼性予測の過程について説明する．

表 20　ある欠測パターンに対応した完全データから求めた予測値.

	真値	予測値
No.	y	$\hat{y}$
1	0	0
⋮	⋮	⋮
n	0	1+

表 21　ある欠測パターンの擬似完全データから求めた予測値.

	真値	予測値					統合された予測値
No.	y	$\hat{y}_I$	$\hat{y}_{II}$	$\hat{y}_{III}$	$\hat{y}_{IV}$	$\hat{y}_V$	$\hat{y}$
1	0	0	1+	0	1+	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	0	1+	0	1+	0	1+	1+

6.1　学習データとテストデータの決定

予測モデルの予測精度評価のために学習データとテストデータを決定する．学習データとテストデータの分割方法は多く提案されているが，本研究では代表的な方法であるホールドアウト法と交差確認法を採用した．この節では分割方法とその後の予測方法について示す．

ホールドアウト法

単純に対象データを学習データ，テストデータの 2 つに分割する方法である．また，この方法はデータ数に依らない．図 10 は本研究のホールドアウト法における $\hat{y}$ の導出過程を示している．38 個の学習データから作成した 1 つの予測モデルを作成する．この予測モデルを 10 個のテストデータに適用することで $\hat{y}$ を求めている．交差確認法と比べると計算量が少ない．

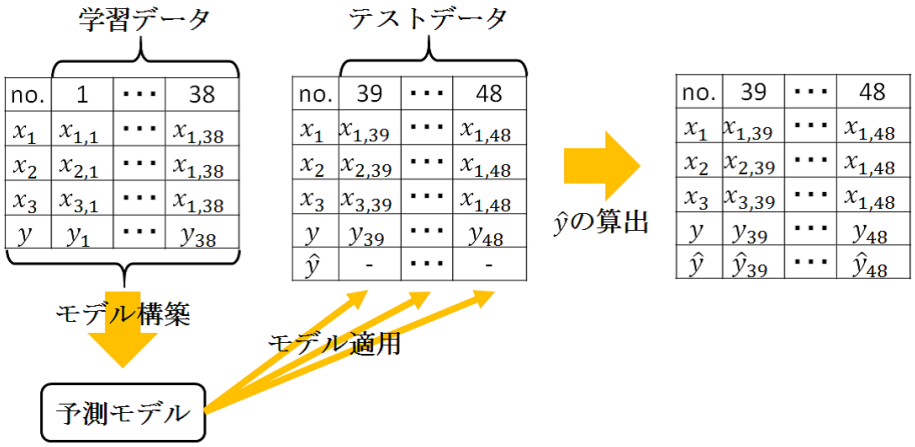

図 10　ホールドアウト法における $\hat{y}$ の導出過程．

交差確認法

対象データ 48 個を 47 個の学習データと 1 個のテストデータに分割し，1 個のテストデータに対して予測モデル適用することで 1 つの $\hat{y}$ を求めている．この導出をデータが重複しないよう 10 回繰り返すことでテストデータ 10 個の $\hat{y}$ を得ている．図 11 は本研究の交差確認法における $\hat{y}$ の導出過程を示している．テストデータはホールドアウト法と同じデータを取り出している．

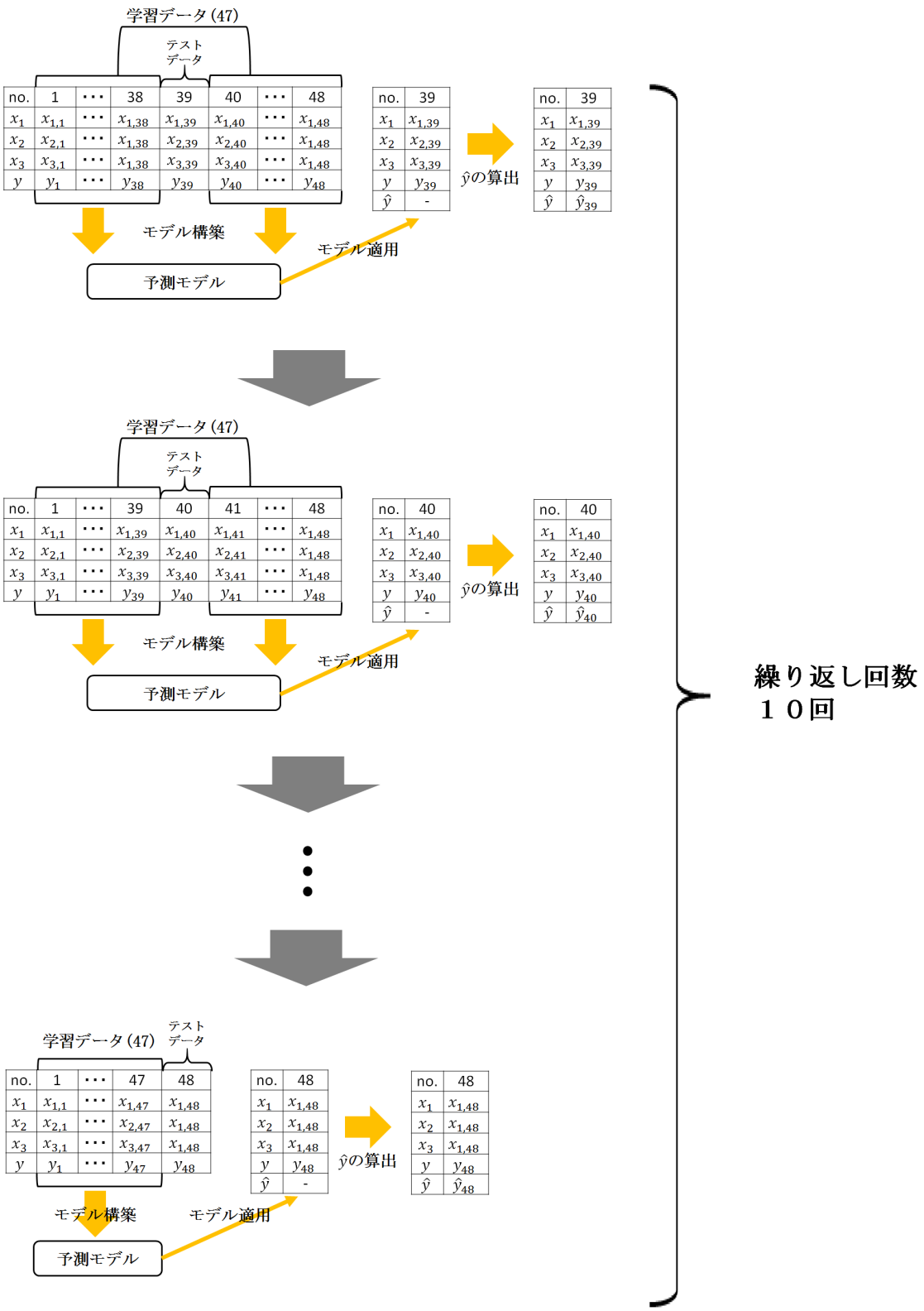

図 11　交差確認法における $\hat{y}$ の導出過程.

テストデータは 48 個のデータの中からランダムに 10 個選んだ．テストデータとなるデータの No. を欠測パターンごとにまとめたものが表 22 である．欠測パターンごとに対象データ No. が異なる理由は，抜きだした 1 つのデータが欠測パターンごとに異なるためである．詳細は第 2.4 節を参照．これらのテストデータに対して欠測パターンごとに予測精度を比較した．

表 22　欠測パターンごとのテストデータ No..

テストデータ No.	対象データ No.		
	欠測パターン		
	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	2	2	2
2	6	6	5
3	9	9	8
4	15	15	14
5	22	22	21
6	28	27	27
7	30	29	29
8	36	36	36
9	42	42	42
10	47	47	47

6.2 信頼性予測過程

この章のはじめに信頼性予測のおおまかな手順のみ説明した．この節では具体的なデータを用いた手順を箇条書きで説明する．

手順 1
: 完全データを学習データとテストデータに分割する．

手順 2
: 学習データを用いて予測モデルを作成する．

手順 3
: 予測モデルを完全データのテストデータに適用し，予測値 $\hat{y}$ を得る．

手順 4
: テストデータの真値と予測値を比較し，正答率を調べる．この正答率が完全データから得たモデルの予測精度である．

手順 5
: 擬似完全データを学習データとテストデータに分割する．

手順 6
: 学習データを用いて予測モデルを作成する．このモデルを擬似完全データのテストデータに適用し，予測値を得る．

手順 7
: 手順 5, 6 を擬似完全データ I〜V まで行う．

手順 8
: 擬似完全データ I〜V の予測値をテストデータ毎に統合することで予測値 $\hat{y}$ を得る．

手順 9
: テストデータの真値と予測値を比較し，正答率を調べる．この正答率が擬似完全データから得たモデルの予測精度である．

手順 10
: 完全データと擬似完全データの予測精度を比較し，共に予測精度が 5 割以上でなおかつ予測精度に差がなければ擬似完全データからでも信頼性予測が可能だとみなせる．

これらの導出過程を欠測パターン (6, 1, 1), (6, 2, 1), (7, 1, 0) に対して行った．

7　分析手法

信頼性予測には予測モデルを用いることを第 6 章にて示した．この章では本研究で用いた 3 種類の予測モデル構築法を説明する．

7.1　Random Forest

Random Forest はアンサンブル学習による機械学習アルゴリズムの 1 つである．決定木を複数構築し，それらを弱識別器として統合することで分類を行う方法である [21]．学習データを $v = (x_1, x_2, x_3, y)$ としたときの Random Forest を表したものが図 12 であり，図中で複数作成されているものが決定木である．決定木とはそれぞれの接点が持つ子のうちどれに進むかを決定する分岐関数が与えられ，葉に最終的な出力結果が対応付けられたものである．接点とは決定木内の分岐点のことであり，子とは接点の下にある接点である．また，葉とは決定木内の最も下にある接点のことである．はじめに v は決定木の根に入力される．根では分割関数の評価結果に応じてどの子にデータが移るかが決まる．この処理を繰り返して葉にデータが到達すると葉に対応づけられた結果が出力される．根とは決定木内の最も上にある接点のことである．

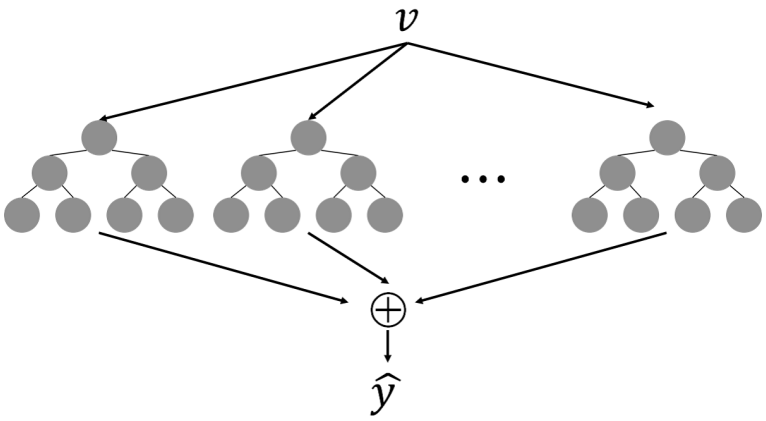

図 12　Random Forest の概要.

本研究では Random Forest を用いるにあたって R のパッケージ ‘randomForest ’ を用いた [22]．このパッケージに沿った Random Forest 法と CART 法について説明する．また，執筆にあたり文献 [23, 24, 25, 26] を参考にした．

CART 法とは樹木構造接近法の 1 つである．CART 法は

1. 前進過程：決定木の成長過程
2. 後退過程：決定木の剪定
3. 最適モデル選択過程：最適な木の決定

の 3 ステップからなる．n をデータ数とし，データ集合 $v = (x_n, y_n)$ と決めるとステップ 1 では目的変数 y_n が K 個のクラス $\{1, \cdots, k\}$ のいずれかの値をとるとする．本研究の場合，クラスは「0」,「1+」であるため

$k=2$ である．このとき接点 t において目的変数が特定のクラス k をとる確率の推定値は

$$\Pr(k|t) = \sum_{n \in t} \mathbf{1}(y_n = k)/N(t) \ , \tag{7.1}$$

である．また，$\mathbf{1}(\cdot)$ は括弧内が真の場合は 1，偽の場合は 0 を返す関数である．接点 t における $\hat{y}_n$ は多数決を用いて

$$\hat{k}(t) = \arg\max_{k} \Pr(k|t) \ , \tag{7.2}$$

である．接点 t の不均一性測度として Gini 係数を用いた．不均一性測度とは回帰問題における残差平方和にあたる．したがってこれは接点 t が y_n の実測値通りに分類できているかを表す指標である．Gini 係数とは

$$r(t) = \sum_{k \neq k'} \Pr(k|t) \Pr(k'|t) = \sum_{1 \leq k \leq K} \Pr(k|t)(1 - \Pr(k|t)) \ , \tag{7.3}$$

である．Gini 係数を用いて表現した

$$R(t) = \Pr(t) r(t) \ , \tag{7.4}$$

はリスクの再代用推定値として接点 t の分岐関数 s_t を形成する．分岐関数 s_t は説明変数 $x_p, (p = 1, 2, 3)$ とするとこの中からランダムに選ばれた説明変数 P 個を分岐基準に組み込む．本研究では $P = 2$ である．こうして選ばれた説明変数を $x_q, (q = 1, 2)$ とすると接点 t で候補となる分岐関数 s_t の集合を S_t と決めた．このとき接点 t 以下に形成する子を t_L, t_R としたとき t に属する個体が分岐関数 s_t を満たせば t_L に送られ，満たさなかった場合は t_R に送られる．分岐関数 s_t に対する分岐測度は

$$R(s_t, t) = R(t) - R(t_L) - R(t_R) \ , \tag{7.5}$$

で求められる．このとき S_t の中で最小の減少量となる分岐関数を

$$s_t^* = \arg\max_{s_t \in S_t} \triangle R(s_t, t) \ , \tag{7.6}$$

と決めた．この s_t^* が接点 t の分岐を決める．

この分岐が実行されるたびに子が次々に作成される．この分岐は「葉内のデータ数 $N(t)$ が l よりも小さくなる」を条件に停止すると決めた．本研究では $l = 1$ とした．通常の CART 法はこのような木の生成の後に木の後退過程と最適モデルの選択過程があるが，本研究の Random Forest では前進過程内の停止条件 1 つのみで決定木と定めた．

ここまで決定木の作成方法について述べた．Random Forest ではこの決定木を複数作成し，各々の決定木から得た $\hat{y}$ を統合することで最終的な予測値を得ている．また，決定木の作成過程で分岐関数を決める際に，組み込む説明変数をランダムに決めるため各決定木の分岐関数は異なる．こうして得られた決定木群が $T_b^{(max)}, (b = 1, \cdots, B)$ である．B は作成された決定木の数であり，本研究では $B = 500$ とした．決定木 $T_b^{(max)}$ の予測値 $\hat{y}$ を返す予測子を $h(x; v_b)$ とし，統合された予測結果を $\hat{f}_{(B)}^{RF}(x)$ とすると

$$\hat{f}_{(B)}^{RF}(x) = \arg\max_{1 \leq k \leq K} \left\{ \sum_{b=1}^{B} \mathbf{1}(h(x; v_b) = k) \right\} \ , \tag{7.7}$$

と書ける．この $\hat{f}_{(B)}^{RF}(x)$ が予測モデルである．この予測モデルにテストデータを適用することで予測値 $\hat{y}$ を得ている．

7.2 ロジスティック回帰

ロジスティック回帰とは説明変数 x が量的変数，目的変数 y が 2 値の質的変数の際に用いられる回帰分析である．本研究の目的変数 y は「0」,「1+」をとる 2 値の質的変数であり，説明変数 $x_p, (p = 1, 2, 3)$ は量的変数であるためロジスティック回帰を用いることができた．この節ではロジスティック回帰の予測モデルの導出について説明する．説明するにあたり文献 [27, 28] を参考にした．

y が 1+ のとき $y = 1$ とし，y が 0 のとき $y = 0$ と決めるとそれぞれの確率を

$$p = \Pr(y = \ 1 \ | x_1, x_2, \cdots, x_p) \ , \tag{7.8}$$

$$1 - p = \Pr(y = 0 | x_1, x_2, \cdots, x_p) \ , \tag{7.9}$$

と表現できる．このときロジスティック回帰の定数項を b_0，偏回帰係数を $b_p, (p = 1, 2, 3)$ とすると

$$p = \frac{\exp(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}{1 + \exp((b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)} \ , \tag{7.10}$$

のように 1 つの式にすることができる．これをロジスティック回帰モデルとする．式 (7.10) をロジット変換することで

$$\log \frac{p}{1 - p} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p \ , \tag{7.11}$$

となる．表記の簡便化のため $\boldsymbol{b^T} = (b_0, b_1, \cdots, b_p)$ とすると

$$\log \frac{p}{1 - p} = \boldsymbol{b^T x} \ , \tag{7.12}$$

となり，式 (7.11) は線形回帰モデルとなった．次にデータ数を n，$i = (1, 2, \cdots, n)$ とすると p, y, i を用いて

$$f(y_i | p_i) = p_i^{y_i} (1 - p_i)^{1 - y_i} \ , \tag{7.13}$$

と表現できる．これは y がベルヌーイ分布に従うためである．この離散型確率分布を使って最尤法によりパラメータ推定を行うことができる．尤度関数は

$$\begin{aligned} L(p_i | y_i) &= \prod_{i=1}^{n} f(y_i | p_i) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1 - y_i} = \prod_{i=1}^{n} \left(\frac{p_i}{1 - p_i} \right)^{y_i} (1 - p_i) \\ &= \sum_{i=1}^{n} y_i \log \frac{p_i}{1 - p_i} + \sum_{i=1}^{n} \log(1 - p_i) \ , \end{aligned} \tag{7.14}$$

である．また，式 (7.14) に式 (7.10), 式 (7.13) を代入し，対数をとることで作成した対数尤度関数は

$$\begin{aligned} \log L(b) &= \sum_{i=1}^{n} y_i \log \left\{ \exp(b^T x_i) \right\} + \sum_{i=1}^{n} \log \frac{1}{1 + \exp(b^T x_i)} \\ &= \sum_{i=1}^{n} y_i b^T x_i - \sum_{i=1}^{n} \log \left\{ 1 + \exp(b^T x_i) \right\} \ , \end{aligned} \tag{7.15}$$

である．式 (7.15) の最大値を求めるために推定したい目的のパラメータ b_p で偏微分して 0 とおき，方程式を解くことで最尤推定量 $\hat{b}_p$ を得ることができる．次に $\hat{b}_p$ を式 (7.10) に代入することで p を得る．この p を用いて $p \leq 0.5$ のとき $\hat{y} = 0$，$p > 0.5$ のとき $\hat{y}$ =1+ と判断した．

上記のロジスティック回帰モデルを対象データの学習データに適用して予測モデルを作成した．この導出は R の関数 glm() を用いた．

7.3 判別分析

いくつかの母集団が混在して観測されたデータがあり，各観測値はこれらの母集団のうち 1 つの母集団に属しているとする．いずれかに属する不明なデータがあり，このデータがどの母集団に属するかを判定する規則を見つけることが判別の問題である．このような問題に対してどの群に属するかを識別するための判別関数 $h(x)$ を作成し，その関数に新たなデータを通すことで得られた関数値が基準値 c に対して $h(x) > c$ の場合であれば第 1 群に，$h(x) \leq c$ であれば第 2 群に所属すると判定する．このような分類が判別分析である．判別分析では判別の成功率が最大になるように判別関数を作成する．本研究の場合，観測値では第 1 群を $y = 0$，第 2 群を $y =1+$ と決めた．また，新たなデータとはテストデータのことである．本研究では尤度に基づいた線形判別の場合について考えていく．この節は文献 [27, 29] を参考にした．

変数 $x_p, (p = 1, 2, 3, 4)$ とすると

x_1: SLOC(Source Lines of Code).
x_2: 計画月数.
x_3: 平均要員数.
x_4: ソフトウェアリリース後 1 ヵ月以内の不具合有無.

と決めた．この変数ごとの係数を $a_p, (p = 1, 2, 3, 4)$ とし，

$$z = a_1x_1 + a_2x_2 + \cdots + a_px_p \ , \tag{7.16}$$

という z 軸を決めた．任意の変数 x_p とそれを第 1 群と第 2 群の平均で表した変数 $\bar{x}^{(1)}$，$\bar{x}^{(2)}$ と a_p をベクトルで表現したものが

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}, \qquad \bar{\boldsymbol{x}}^{(\boldsymbol{1})} = \begin{pmatrix} \bar{x}_1^{(1)} \\ \bar{x}_2^{(1)} \\ \vdots \\ \bar{x}_p^{(1)} \end{pmatrix}, \qquad \bar{\boldsymbol{x}}^{(\boldsymbol{2})} = \begin{pmatrix} \bar{x}_1^{(2)} \\ \bar{x}_2^{(2)} \\ \vdots \\ \bar{x}_p^{(2)} \end{pmatrix}, \qquad \boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix},$$

である．このベクトルを用いて標本分散共分散行列は

$$\bar{\boldsymbol{S}}^{(\boldsymbol{1})} = \begin{pmatrix} s_{11}^{(1)} & s_{12}^{(1)} & \cdots & s_{1p}^{(1)} \\ s_{21}^{(1)} & s_{22}^{(1)} & \cdots & s_{2p}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1}^{(1)} & s_{p2}^{(1)} & \cdots & s_{pp}^{(1)} \end{pmatrix}, \qquad \bar{\boldsymbol{S}}^{(\boldsymbol{2})} = \begin{pmatrix} s_{11}^{(2)} & s_{12}^{(2)} & \cdots & s_{1p}^{(2)} \\ s_{21}^{(2)} & s_{22}^{(2)} & \cdots & s_{2p}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1}^{(2)} & s_{p2}^{(2)} & \cdots & s_{pp}^{(2)} \end{pmatrix},$$

とした．また，$\bar{\boldsymbol{x}}^{(\boldsymbol{1})}$ のデータ数を n_1，$\bar{\boldsymbol{x}}^{(\boldsymbol{2})}$ をデータ数 n_2 とすると $\bar{\boldsymbol{x}}^{(\boldsymbol{1})}$ と $\bar{\boldsymbol{x}}^{(\boldsymbol{2})}$ の共通の分散共分散行列を

$$\boldsymbol{S} = \frac{(n_1 - 1)\boldsymbol{S}^{(\boldsymbol{1})} + (n_2 - 1)\boldsymbol{S}^{(\boldsymbol{2})}}{n_1 + n_2 - 2} \ , \tag{7.17}$$

と表せる．$\boldsymbol{S}$ が式 (7.17) により求まるとき

$$f_k(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(k)}, \boldsymbol{S}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{S}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\boldsymbol{k})})^T\boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\boldsymbol{k})})\right\} (k = 1,\ 2) \ , \tag{7.18}$$

のように各群のデータが多変量正規分布に従っているとする． 2 つの確率密度の値 $f_1(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(1)}, \boldsymbol{S}), f_2(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})$ を計算し, 値が大きい確率密度関数をとった群へ判別する．第 1 群に判別するとき, $f_1(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(1)}, \boldsymbol{S}), f_2(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S})$

の対数をとって比較すると

$$\begin{aligned}
\log f_1(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(1)}, \boldsymbol{S}) &> \log f_2(\boldsymbol{x}|\bar{\boldsymbol{x}}^{(2)}, \boldsymbol{S}) \\
-\frac{p}{2} - \frac{1}{2}\log|\boldsymbol{S}| - \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})}) &> -\frac{p}{2} - \frac{1}{2}\log|\boldsymbol{S}| - \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})}) \\
-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})}) &> -\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})}) \\
r(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})}) &> -(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})}) \ , \qquad (7.19)
\end{aligned}$$

式 (7.19) を等式変形し，

$$\begin{aligned}
h(x) &= -(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{1})}) + (\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})})^T \boldsymbol{S}^{-1}(\boldsymbol{x} - \bar{\boldsymbol{x}}^{(\mathbf{2})}) > 0 \\
&= (\bar{\boldsymbol{x}}^{(\mathbf{1})} - \bar{\boldsymbol{x}}^{(\mathbf{2})})^T \boldsymbol{S}^{-1} \left\{ \boldsymbol{x} - \frac{1}{2}\bar{\boldsymbol{x}}^{(\mathbf{1})} + \bar{\boldsymbol{x}}^{(\mathbf{2})} \right\} > 0 \ , \qquad (7.20)
\end{aligned}$$

となる．したがって，$h(x) > 0$ のときは $\boldsymbol{x}$ を第 1 群に判別し，$h(x) \leq 0$ のとき第 2 群に判別する規則が導かれた．式 (7.20) を予測モデルとしてテストデータの $\hat{y}$ を求めた．

8　予測結果と考察

この章では 3 つの予測モデルから得られた予測結果を示し，考察を述べる．また，擬似完全データから得られた予測結果は欠測パターンごとに統合された結果のみを示す．

8.1　Random Forest

Random Forest を予測モデルとしたときの予測値をデータ分割法ごとにまとめたものが表 23, 24 である．表 23 から, 完全データの予測精度は全ての欠測パターンにおいて 6 割を超えていることから信頼性予測への若干の有用性は認められた．しかし，擬似完全データの予測精度は全ての欠測パターンで 4 割であることから信頼性予測には使えないことが分かった．以上のように完全データと擬似完全データの予測精度には差がでてしまったため，ホールドアウト法によって分割された学習データを組み込んだ Random Forest による信頼性予測は不完全データに対しては有用な予測ではないと分かった．

一方で，表 24 から完全データの予測精度は全ての欠測パターンにおいて 4 割以下であることから信頼性予測の精度が低いことが分かった．また，擬似完全データの予測精度は全ての欠測パターンで 5 割以下であることから同様に信頼性予測には使えないことが分かった．

以上のように完全データと擬似完全データの予測精度には若干の差がでてしまったことに加えて，予測精度が低い．したがって，交差確認法によって分割された学習データを組み込んだ Random Forest による信頼性予測は有用な予測ではないと分かった．

表 23　ホールドアウト法による Random Forest の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	1+	1+	1+	0	0	0
2	1+	1+	1+	1+	0	0	0
3	0	1+	1+	1+	1+	1+	1+
4	1+	1+	1+	1+	1+	0	0
5	0	1+	1+	1+	1+	1+	1+
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	1+	1+	1+	1+	0	0
10	0	1+	1+	1+	1+	1+	1+
予測精度		6	6	6	4	4	4

表 24　交差確認法による Random Forest の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	0	1+	1+	1+	0	0
2	1+	0	0	0	0	0	0
3	0	1+	1+	1+	1+	1+	1+
4	1+	0	0	0	1+	0	0
5	0	1+	1+	1+	1+	1+	1+
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	1+	1+	1+	1+	0	0
10	0	1+	1+	1+	1+	1+	1+
予測精度		3	4	4	5	4	4

8.2 ロジスティック回帰

ロジスティック回帰を予測モデルとしたときの予測値をデータ分割法ごとにまとめたものが表 25, 26 である．表 25 から, 完全データの予測精度は全ての欠測パターンにおいて 4 割以下であることから信頼性予測の精度が低いことが分かった．また，擬似完全データの予測精度は欠測パターン (7, 1, 0) においてのみ 6 割であり，他の欠測パターンでは 5 割未満であることから同様に信頼性予測には使えないことが分かる．以上のように完全データと擬似完全データの予測精度は欠測パターン (7, 1, 0) 以外はほぼ同様であったが，予測精度がほとんどの欠測パターンにおいて低い．したがって，ホールドアウト法によって分割された学習データを組み込んだロジスティック回帰による信頼性予測が有用な予測ではないと分かった．

一方で，表 26 から完全データの予測精度は全ての欠測パターンにおいて 4 割以下であることから信頼性予測の精度が低いことが分かった．また，擬似完全データの予測精度は全ての欠測パターンも同様に 4 割以下であることから同様に信頼性予測には使えないことが分かった．しかし，欠測パターンごとの予測精度は等しく，差がないことが分かった．

以上のように完全データと擬似完全データの予測精度には差が見受けられなかったものの，予測精度が低い．したがって，交差確認法によって分割された学習データを組み込んだロジスティック回帰による信頼性予測は有用な予測ではないと分かった．

表 25　ホールドアウト法によるロジスティック回帰の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	1+	1+	1+	1+	1+	1+
2	1+	0	0	0	0	0	0
3	0	1+	1+	1+	1+	1+	1+
4	1+	1+	1+	1+	1+	1+	1+
5	0	1+	0	0	1+	0	1+
6	0	1+	1+	1+	1+	1+	1+
7	0	1+	1+	1+	1+	1+	0
8	0	1+	1+	1+	1+	1+	0
9	0	0	0	0	0	0	0
10	0	1+	1+	1+	1+	1+	0
予測精度		3	4	4	3	4	6

表 26　交差確認法によるロジスティック回帰の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	1+	1+	1+	1+	1+	1+
2	1+	0	0	0	0	1+	0
3	0	1+	1+	1+	1+	1+	1+
4	1+	1+	1+	1+	1+	1+	1+
5	0	0	0	0	0	0	1+
6	0	1+	1+	1+	1+	1+	1+
7	0	1+	1+	1+	1+	1+	1+
8	0	1+	1+	1+	1+	1+	1+
9	0	1+	0	0	1+	1+	0
10	0	1+	1+	1+	1+	1+	0
予測精度		3	4	4	3	4	4

8.3 判別分析

判別分析を予測モデルとしたときの予測値をデータ分割法ごとにまとめたものが表 27, 28 である．表 27 から, 完全データの予測精度は全ての欠測パターンにおいて 7 割であることから予測精度は有用性があると判断できる．しかし予測値が全て 0 であることから有効な予測であるかどうか疑問が残る結果となった．また，擬似完全データの予測精度は欠測パターン (6, 1, 1) においてのみ 7 割であり，他の欠測パターンでは 5 割以下であることから欠測パターンによっては予測精度が十分でない．

以上のように完全データと擬似完全データの予測精度には若干の差がでてしまったことに加えて，予測精度が欠測パターンによって予測精度が大きく異なる．したがって，ホールドアウト法によって分割された学習データを組み込んだ判別分析による信頼性予測は不完全データに対しては有用な予測ではないと分かった．

一方で，表 28 から完全データの予測精度は全ての欠測パターンにおいて 6 割以上であることから予測精度の有用性があると判断した．また，擬似完全データの予測精度は全ての欠測パターンで等しく，予測精度は十分である．以上のことから，完全データと擬似完全データの予測精度には差が見受けられないことに加え，予測精度も 5 割を超えていることから有用性はあると判断した．したがって，交差確認法によって分割された学習データを組み込んだ判別分析による信頼性予測は完全データ，不完全データに対して有用な予測であることが分かった．

表 27　ホールドアウト法による判別分析の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	0	0	0	0	0	0
2	1+	0	0	0	1+	0	1+
3	0	0	0	0	0	0	0
4	1+	0	0	0	0	0	0
5	0	0	0	0	0	1+	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1+
8	0	0	0	0	0	0	1+
9	0	0	0	0	1+	1+	1+
10	0	0	0	0	0	0	1+
予測精度		7	7	7	7	5	4

表 28　交差確認法による判別分析の予測値と予測精度.

テストデータ		欠測パターン					
		完全データ			擬似完全データ		
No.	真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
1	1+	0	0	0	0	0	0
2	1+	1+	1+	1+	1+	0	1+
3	0	0	0	0	0	0	0
4	1+	0	0	0	0	0	0
5	0	1+	1+	1+	1+	1+	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	1+	1+	0	0	1+
10	0	0	0	0	0	0	1+
予測精度		7	6	6	7	6	6

8.4 考察

信頼性予測にて用いた学習法とテストデータの分割法はホールドアウト法と交差確認法を用いたが，通常ならば学習データの数が多い交差確認法の予測精度がより良くなることが予想される．この仮定が正しいかどうか予測モデルごとに確認した．完全データの Random Forest ではホールドアウト法の予測精度が良く，擬似完全データでは交差確認法の予測精度が若干良い．完全データのロジスティック回帰では交差確認法の予測精度が良く，擬似完全データではホールドアウト法の予測精度が若干良い．また，完全データの判別分析ではホールドアウト法の予測精度が若干良く，擬似完全データでは交差確認法の予測精度が若干良い．したがってホールドアウト法と交差確認法による予測精度の大きい差は生じないことがわかった．

表 28 から，交差確認法による判別分析が全ての欠測パターンにおいて予測精度が等しく，6 割を超えていることから予測モデルとして優れているという結果となった．予測精度は等しいものの，予測値が等しいのは欠測パターン (6, 1, 1) のみである．したがって，予測モデルが全く同じ予測値を出力したわけではない．

真値ごとの予測精度をまとめたものが表 29 である．真値は 1+ が 3 個, 0 が 7 個であり，この表が示す通り不具合が 0 の予測は当てはまりが良いが，1+ の検出力は低いことが分かる．

表 29　交差確認法による判別分析の真値ごとの予測精度.

テスト	欠測パターン					
データ	完全データ			擬似完全データ		
真値	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
0	6/7	5/7	5/7	6/7	6/7	5/7
1+	1/3	1/3	1/3	1/3	0/3	1/3

次に欠測パターンごとの生産者危険と消費者危険の予測精度について記述する．生産者危険とは抜き取り検査において合格にすべきものを不合格としてしまう確率のことである．消費者危険とは不合格にすべきものを合格にしてしまう確率のことである．本研究の場合，生産者危険とは目的変数が「真値 0 のとき予測値 0.」,「真値 0 のとき予測値 1+.」,「真値 1+ のとき予測値 1+.」となる場合のことを表す．一方で，消費者危険とは「真値 1+ のとき予測値 0」となることである．欠測パターン (6, 1, 1) に対応する完全データに判別分析を適用した予測精度をまとめたものを表 30 とする．さきほどの考えに照らすと，生産者危険が $(6+1+1)/10 = 8/10$ で，消費者危険が $2/10$ である．

表 30　欠測パターン (6, 1, 1) に対応する完全データに判別分析を適用した予測精度.

		真値	
		0	1+
予測値	0	6	2
	1+	1	1

生産者危険と消費者危険を欠測パターンごとにまとめたものが表 31 である．この表から，生産者危険は完全データと擬似完全データの欠測パターン (6, 2, 1) において若干の誤差が生じてしまったことが分かる．

表 31　交差確認法による判別分析の生産者危険．

	完全データ			擬似完全データ		
欠測パターン	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)	(6, 1, 1)	(6, 2, 1)	(7, 1, 0)
生産者危険	8/10	8/10	8/10	8/10	7/10	8/10

9　おわりに

本研究では欠測値を含む不完全データに対して完全データと同じように信頼性予測を行えることを示すことが目的である．そのために，元データから完全データと，完全データとほぼ同様の不完全データを作成した．次に多重代入法の 1 つである EMB アルゴリズムを用いて不完全データの欠測値を代入し，擬似完全データを作成した．こうして得られた完全データと擬似完全データの各データを，学習データとテストデータに分解した．分解方法はホールドアウト法と交差確認法を採用した．また，学習データは予測モデルの構築に用いられ，テストデータは予測モデルの予測精度を測るために利用された．その結果，予測モデルの 1 つである判別分析を用いて交差確認法により得られた完全データと擬似完全データのテストデータを予測したところ，予測精度に差がないことが見受けられた．したがって，交差確認法により分割されたデータを用いて判別分析を適用し，予測精度を求める場合には完全データと不完全データ間に精度の差は生じないと判断できることがわかった．

完全データと擬似完全データ間で予測精度に差が生じないようにするために，欠測率と欠測パターンという考え方を適用した．また，予測精度向上のために複数の予測モデルを適用して有用なモデルを模索した．このように本研究では予測精度の一致と向上が研究の重要課題であった．２つの課題に対して検討の余地が残るものをこの章で述べ，本論文の結びにしたいと思う．

説明変数の数

今回は 3 つの説明変数で目的変数を予測した．変数の数が多い場合，欠測パターンを考える際に今回の (6, 1, 1), (6, 2, 1), (7, 1, 0) より多くなる可能性はある．本研究では全ての欠測パターンで予測精度が一致する予測モデルの構築に成功したが，欠測パターンが増加するほど完全データと擬似完全データから得られた予測精度を一致させることは難しくなる．したがって説明変数の増加は好まれない面がある．一方で，元データは変数が 611 個あるため説明変数候補は数多く存在する．説明変数を多くすれば考慮できる情報も多くなるため予測精度向上を期待できるが，変数を増加することで欠測パターンの増加も同時に起こる．最適な説明変数の数を考える必要がある．

擬似完全データの M

多重代入法において欠測値代入を行う回数 M を本研究では $M = 5$ とした．このように決めた理由はデータの取り扱いの簡便化と再現性の保持である．EMB アルゴリズムでは変数内でノンパラメトリックブートストラップが行われ，Random Forest では決定木を数百個作成する．このようにデータの復元抽出や副標本の作製が数多く行われていることから，データの取り扱いを容易にするために欠測値代入の回数は最小限に抑えた．また，分析には R 言語を用いていることから，計算が正しいかどうかを確認するためのデータを保持した．このように保持したデータは修士論文の付録とすることで再現性の確保の役目も果たしている．この M の数が膨大になってしまうと再現性の確保が難しいことから M は最小限に抑えた．しかし，M を増加させることで予測精度の向上も見込めることから M の増加も課題として残った．

他の欠測値代入の適用

EMB アルゴリズムを用いて作成された擬似完全データと完全データに対して行った 3 つの予測モデルにおいて，予測精度に大きな差はないことから本研究では欠測値代入法として採用した．したがって，擬似完全データは 1 つの欠測代入法からのみ得られている．しかし欠測値代入法は他にも考案されている．

参考文献

[1] 情報処理用語-基本用語 JIS 原案作成委員会：情報処理用語 - 基本用語　(JIS X 0001 - 1994), 日本工業規格, (1994).

[2] IPA/SEC：ソフトウェア開発データ白書 2012-2013，IPA/SEC (2012).

[3] Mint (経営情報研究会)：図解でわかるソフトウェア開発の全て, 日本実業出版社　(2012).

[4] 野下貴弘，森田貴之: 「プロジェクトにおける品質成功度評価モデルの作成と考察」，法政大学卒業論文 (2013).

[5] Morita, T., Esaki, K., Kimura, M.： “A Note on Modeling of Quality Evaluation Based on Large Data Sets in Software Development Projects ”, *14th APIEMS 2013 Conference Proceedings* (2013).

[6] G. E. P. Box, D. R. Cox： “An Analysis of Transformations ”, *Journal of the Royal Statistical Society*, Vol. 26, No. 2, pp. 211-252 (1964).

[7] S. S. Shapiro, M. B. Wilk： “An Analysis of Variance Test for Normality (Complete Samples) ”, *Biometrika*, Vol. 52, No. 3/4, pp. 591-611 (1965).

[8] R. J. Little, R. D’Agostino, M. L. Cohen, et al.： “The prevention and treatment of missing data in clinical trials ”, *New England Journal of Medicine* , 367, pp. 1355 - 1360 (2012).

[9] 日本製薬工業協会，医薬品評価委員会 データサイエンス部会，2013 年度タスクフォース 2：「臨床試験の欠測データの取り扱いに関する最近の展開と今後の課題について- NAS レポート，EMA ガイドライン，estimand と解析方法の概説 -」，製薬協 (2013).

[10] D. B. Rubin, R. J. A. Little： *Statistical Analysis with Missing Data, Second Edition*, *Wiley-—Interscience*, New Jersey (2002).

[11] 高橋将宜，伊藤孝之：「経済調査における売上高の欠測値補定値方法について～多重代入法による精度の評価～」，統計研究彙報，Vol. 70, pp. 19-86 (2013).

[12] 高橋将宜，伊藤孝之：「様々な多重代入アルゴリズムの比較～大規模経済系データを用いた分析～」，統計研究彙報，Vol. 71, pp. 39-82 (2014).

[13] 野間久史：欠測データの統計解析：理論と応用，統計数理研究所，2014 年度公開講座資料 (2014).

[14] D. B. Rubin： *Multiple Imputation for Nonresponse in Surveys*, *John Wiley & Sons, Inc*, Hoboken (2004).

[15] S. Van Buuren, J. P. L. Brand, C. G. M. Groothuis-Oudshoorn, D. B. Rubin, : “Fully conditional specification in multivariate imputation ”, *Journal of Statistical Computation and Simulation*, Vol. 76, No. 12, pp. 1049-1064 (2006).

[16] J.Honaker, G.King, “What to Do about Missing Values in Time-Series Cross-Section Data”, *American Journal of Political Science*, Vol. 54, No. 2, pp. 561-581 (2010).

[17] G. King, J. Honaker, A. Joseph, K. Scheve, “Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation ”,*American Political Science Review*, Vol. 95, No. 1, pp. 49-69 (2001).

[18] J. Honaker, G. King, M. Blackwell： “Amelia II: A Program for Missing Data”, *Journal of Statistical Software*, Vol. 45, No. 7, pp. 1-42 (2011).

[19] 小西貞則，越智義道，大森裕浩：計算機統計学の方法 - ブートストラップ・EM アルゴリズム・MCMC，朝倉出版 (2008).

[20] 渡辺美智子，山口和範：EM アルゴリズムと不完全データの諸問題，多賀出版 (2000).
[21] L. Breiman. : “Random Forests”, *Machine Learning*, 45, 1, pp. 5-32 (2001).
[22] A. Liaw, M. Wiener : “Package ‘randomForest ”, *CRAN* , Ver. 4.6 -10 (2014).
[23] 杉本知之，下川敏雄，後藤昌司：樹木構造接近法（R で学ぶデータサイエンス），共立出版 (2013).
[24] 下川敏雄，杉本知之，後藤昌司：「樹木構造接近法と最近の発展」，計算機統計学，Vol. 18, No. 2, pp. 123-164 (2007).
[25] 波部　斉：「ランダムフォレスト」，情報処理学会研究報告，Vol. 20, No. 31, pp. 1-8 (2012).
[26] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone : *Classification and Regression Trees*, *Chapman and Hall / CRC* (1984).
[27] 中村永友：多次元データ解析法（R で学ぶデータサイエンス），共立出版，(2009).
[28] 丹後俊郎，山岡和枝，高木晴良：ロジスティック回帰分析-SAS を利用した統計解析の実態-，朝倉出版 (1996).
[29] 小西貞則，本多正幸：「判別分析における誤判別率推定とブートストラップ法, Vol. 21, No. 2, pp. 67-100 (1992).

謝辞

本論文を結ぶにあたり，本研究の遂行に際してご指導，ご協力いただきました多くの方々に感謝の意を表します．法政大学理工学研究科の木村光宏教授には本研究と本論文に留まらず，修士課程で執筆した研究論文に関しても終始多大なご指導ご鞭撻を頂きました．２年間という短い期間ではありましたが研究に取り組む真摯さや配慮の細かさなど数えきれないほど多くのことを学ばさせていただきました．心より深く感謝申し上げます．独立行政法人情報処理推進機構の秋田君夫氏，三縄俊信氏，山下博之氏には本研究を行うにあたり貴重なデータを提供していただきました．心より御礼申し上げます．法政大学理工学研究科の江崎和博準准教授には修士論文の副査やデータの取り扱いについて多くのアドバイスをいただきました．心より感謝いたします．経営数理工学研究室の亀岡若菜氏，古賀裕紀氏とは同室で研究することが多く，ともに勉学に励み，励ましあいながら研究を進めることができました．感謝の意を表します．信頼性工学研究室の新井雄一朗氏，太田修平氏からは研究論文執筆など多くの場面で助言をいただきました．厚く御礼申し上げます．同研究室の影山孝夫氏，後藤一成氏，秦直道氏とは研究に関して多くの議論を交わすことで切磋琢磨し，本研究をより良いものにすることができました．心より感謝いたします．同研究室の緒先輩方には研究を進めるにあたり多大なご協力を頂きました．感謝の意を表します．さらに，紙面には書ききれない多くの方々に研究生活は支えられ，その支えのおかげで本論文を完成させることができました．数多くの無礼を詫びるとともに，感謝の意を表します．最後に，自分勝手な小生に対して日頃から惜しみない支援と不自由ない大学生活を提供してくださった家族に心より感謝いたします．

著者の文献リスト

[1] Morita, T. , Esaki, K. , Kimura, M. : “A Note on Modeling of Quality Evaluation Based on Large Data Sets in Software Development Projects ”, *APIEMS 2013 Conference Proceedings* (2013).
[2] Morita, T. , Kimura, M. : “A Robustness Analysis of Imputation Method for Software Development Project Data: Missing Value Treatment for Software Quality Prediction ”, *International Journal of Software Engineering and Its Applications*, （印刷中）.

[3] Morita, T. , Kimura, M. : "A Fundamental Study on Missing Value Treatment for Software Quality Prediction", *Advanced Science and Technology Letters*, Vol. 67, pp. 70-73 (2014).

付録 A

第 2.1 節での Shapiro-Wilk 検定で用いた検定表をここに記載する.

表 32　標本数 n と Shapiro-Wilk 検定から得た検討統計量 W に対応する P 値の表.

	P								
n	0·01	0·02	0·05	0·10	0·50	0·90	0·95	0·98	0·99
3	0·753	0·756	0·767	0·789	0·959	0·998	0·999	1·000	1·000
4	·687	·707	·748	·792	·935	·987	·992	·996	·997
5	·686	·715	·762	·806	·927	·979	·986	·991	·993
6	0·713	0·743	0·788	0·826	0·927	0·974	0·981	0·986	0·989
7	·730	·760	·803	·838	·928	·972	·979	·985	·988
8	·749	·778	·818	·851	·932	·972	·978	·984	·987
9	·764	·791	·829	·859	·935	·972	·978	·984	·986
10	·781	·806	·842	·869	·938	·972	·978	·983	·986
11	0·792	0·817	0·850	0·876	0·940	0·973	0·979	0·984	0·986
12	·805	·828	·859	·883	·943	·973	·979	·984	·986
13	·814	·837	·866	·889	·945	·974	·979	·984	·986
14	·825	·846	·874	·895	·947	·975	·980	·984	·986
15	·835	·855	·881	·901	·950	·975	·980	·984	·987
16	0·844	0·863	0·887	0·906	0·952	0·976	0·981	0·985	0·987
17	·851	·869	·892	·910	·954	·977	·981	·985	·987
18	·858	·874	·897	·914	·956	·978	·982	·986	·988
19	·863	·879	·901	·917	·957	·978	·982	·986	·988
20	·868	·884	·905	·920	·959	·979	·983	·986	·988
21	0·873	0·888	0·908	0·923	0·960	0·980	0·983	0·987	0·989
22	·878	·892	·911	·926	·961	·980	·984	·987	·989
23	·881	·895	·914	·928	·962	·981	·984	·987	·989
24	·884	·898	·916	·930	·963	·981	·984	·987	·989
25	·888	·901	·918	·931	·964	·981	·985	·988	·989
26	0·891	0·904	0·920	0·933	0·965	0·982	0·985	0·988	0·989
27	·894	·906	·923	·935	·965	·982	·985	·988	·990
28	·896	·908	·924	·936	·966	·982	·985	·988	·990
29	·898	·910	·926	·937	·966	·982	·985	·988	·990
30	·900	·912	·927	·939	·967	·983	·985	·988	·900
31	0·902	0·914	0·929	0·940	0·967	0·983	0·986	0·988	0·990
32	·904	·915	·930	·941	·968	·983	·986	·988	·990
33	·906	·917	·931	·942	·968	·983	·986	·989	·990
34	·908	·919	·933	·943	·969	·983	·986	·989	·990
35	·910	·920	·934	·944	·969	·984	·986	·989	·990
36	0·912	0·922	0·935	0·945	0·970	0·984	0·986	0·989	0·990
37	·914	·924	·936	·946	·970	·984	·987	·989	·990
38	·916	·925	·938	·947	·971	·984	·987	·989	·990
39	·917	·927	·939	·948	·971	·984	·987	·989	·991
40	·919	·928	·940	·949	·972	·985	·987	·989	·991
41	0·920	0·929	0·941	0·950	0·972	0·985	0·987	0·989	0·991
42	·922	·930	·942	·951	·972	·985	·987	·989	·991
43	·923	·932	·943	·951	·973	·985	·987	·990	·991
44	·924	·933	·944	·952	·973	·985	·987	·990	·991
45	·926	·934	·945	·953	·973	·985	·988	·990	·991
46	0·927	0·935	0·945	0·953	0·974	0·985	0·988	0·990	0·991
47	·928	·936	·946	·954	·974	·985	·988	·990	·991
48	·929	·937	·947	·954	·974	·985	·988	·990	·991
49	·929	·937	·947	·955	·974	·985	·988	·990	·991
50	·930	·938	·947	·955	·974	·985	·988	·990	·991

付録 B

第 5.2 節の対象データのうち完全データを付録 B としてここに列挙する.

表 33　欠測パターン (6, 1, 1) に対応した完全データ.

No.	$x_1^{(6,1,1)}$	$x_2^{(6,1,1)}$	$x_3^{(6,1,1)}$	$y^{(6,1,1)}$	No.	$x_1^{(6,1,1)}$	$x_2^{(6,1,1)}$	$x_3^{(6,1,1)}$	$y^{(6,1,1)}$
1	19.51	1.65	0.87	1+	25	29.02	3.11	2.52	1+
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	19.84	1.56	0.96	1+	27	13.70	1.89	2.55	0
4	20.56	2.29	2.56	0	28	17.14	1.75	1.16	0
5	26.04	3.28	3.15	1+	29	14.50	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.20	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.20	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.30	1.75	0	35	14.53	1.45	0.00	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.20	0	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	1+
15	13.32	1.49	-0.10	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.10	1+	40	16.64	1.76	0.41	1+
17	15.75	2.15	-0.67	0	41	16.74	1.60	0.00	0
18	25.05	2.71	0.72	0	42	19.12	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	0	44	21.37	3.56	2.73	1+
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	21.96	2.79	1.75	1+
24	13.29	2.71	-2.45	1+	48	20.29	1.43	2.82	1+

表 34　欠測パターン (6, 2, 1) に対応した完全データ.

No.	$x_1^{(6,2,1)}$	$x_2^{(6,2,1)}$	$x_3^{(6,2,1)}$	$y^{(6,2,1)}$	No.	$x_1^{(6,2,1)}$	$x_2^{(6,2,1)}$	$x_3^{(6,2,1)}$	$y^{(6,2,1)}$
1	19.51	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.70	1.89	2.55	0
3	19.84	1.56	0.96	1+	27	17.14	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.50	2.52	1.49	0
5	26.04	3.28	3.15	1+	29	18.85	1.74	2.20	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.20	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	1+
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	15.49	1.39	0.72	0
11	27.36	2.30	1.75	0	35	14.53	1.45	0.00	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.20	0	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	1+
15	13.32	1.49	-0.10	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.10	1+	40	16.64	1.76	0.41	1+
17	15.75	2.15	-0.67	0	41	16.74	1.60	0.00	0
18	25.05	2.71	0.72	0	42	19.12	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	0	44	21.37	3.56	2.73	1+
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	21.96	2.79	1.75	1+
24	13.29	2.71	-2.45	1+	48	20.29	1.43	2.82	1+

表 35　欠測パターン (7, 1, 0) に対応した完全データ.

No.	$x_1^{(7,1,0)}$	$x_2^{(7,1,0)}$	$x_3^{(7,1,0)}$	$y^{(7,1,0)}$	No.	$x_1^{(7,1,0)}$	$x_2^{(7,1,0)}$	$x_3^{(7,1,0)}$	$y^{(7,1,0)}$
1	19.51	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.70	1.89	2.55	0
3	19.84	1.56	0.96	1+	27	17.14	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.50	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.20	0
6	33.08	4.05	4.33	1+	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	0	31	17.20	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	1+
9	21.59	3.75	-0.67	1+	33	15.43	1.15	1.27	1+
10	27.36	2.30	1.75	0	34	15.49	1.39	0.72	0
11	18.86	1.96	1.75	1+	35	14.53	1.45	0.00	1+
12	12.12	0.72	3.20	0	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	0	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.10	0	38	19.71	1.96	-0.67	1+
15	20.69	2.71	0.10	1+	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	0	40	16.64	1.76	0.41	1+
17	25.05	2.71	0.72	0	41	16.74	1.60	0.00	0
18	25.57	3.11	3.68	1+	42	19.12	2.74	1.23	1+
19	15.26	2.46	0.96	0	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	1+
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	0	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	1+	47	21.96	2.79	1.75	1+
24	29.02	3.11	2.52	1+	48	20.29	1.43	2.82	1+

付録 C

第 5.2 節の対象データのうち擬似完全データを付録 C としてここに列挙する.

表 36　欠測パターン (6, 1, 1) に従った擬似完全データ I.

No.	$x^{(6,1,1)}_{(1,\mathrm{I})}$	$x^{(6,1,1)}_{(2,\mathrm{I})}$	$x^{(6,1,1)}_{(3,\mathrm{I})}$	$y^{(6,1,1)}_{\mathrm{I}}$	No.	$x^{(6,1,1)}_{(1,\mathrm{I})}$	$x^{(6,1,1)}_{(2,\mathrm{I})}$	$x^{(6,1,1)}_{(3,\mathrm{I})}$	$y^{(6,1,1)}_{\mathrm{I}}$
1	19.47	1.65	0.87	1+	25	25.15	3.11	2.52	0
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	17.66	1.56	0.96	1+	27	13.7	1.89	2.55	0
4	20.56	2.29	2.56	0	28	14.61	1.75	1.16	0
5	26.04	3.77	3.88	0	29	14.5	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.2	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.2	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	22.82	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	22.36	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 37　欠測パターン (6, 1, 1) に従った擬似完全データ II.

No.	$x^{(6,1,1)}_{(1,\mathrm{II})}$	$x^{(6,1,1)}_{(2,\mathrm{II})}$	$x^{(6,1,1)}_{(3,\mathrm{II})}$	$y^{(6,1,1)}_{\mathrm{II}}$	No.	$x^{(6,1,1)}_{(1,\mathrm{II})}$	$x^{(6,1,1)}_{(2,\mathrm{II})}$	$x^{(6,1,1)}_{(3,\mathrm{II})}$	$y^{(6,1,1)}_{\mathrm{II}}$
1	21.25	1.65	0.87	1+	25	26.6	3.11	2.52	0
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	17.43	1.56	0.96	1+	27	13.7	1.89	2.55	0
4	20.56	2.29	2.56	0	28	17.99	1.75	1.16	0
5	26.04	3.9	3.9	0	29	14.5	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.2	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.2	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	19.43	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	18.56	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 38　欠測パターン (6, 1, 1) に従った擬似完全データ III.

No.	$x^{(6,1,1)}_{(1,\mathrm{III})}$	$x^{(6,1,1)}_{(2,\mathrm{III})}$	$x^{(6,1,1)}_{(3,\mathrm{III})}$	$y^{(6,1,1)}_{\mathrm{III}}$	No.	$x^{(6,1,1)}_{(1,\mathrm{III})}$	$x^{(6,1,1)}_{(2,\mathrm{III})}$	$x^{(6,1,1)}_{(3,\mathrm{III})}$	$y^{(6,1,1)}_{\mathrm{III}}$
1	17.15	1.65	0.87	1+	25	23.06	3.11	2.52	0
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	18.09	1.56	0.96	1+	27	13.7	1.89	2.55	0
4	20.56	2.29	2.56	0	28	13.72	1.75	1.16	0
5	26.04	2.31	0.64	0	29	14.5	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.2	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.2	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	24.4	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	23.36	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 39　欠測パターン (6, 1, 1) に従った擬似完全データ IV.

No.	$x^{(6,1,1)}_{(1,\mathrm{IV})}$	$x^{(6,1,1)}_{(2,\mathrm{IV})}$	$x^{(6,1,1)}_{(3,\mathrm{IV})}$	$y^{(6,1,1)}_{\mathrm{IV}}$	No.	$x^{(6,1,1)}_{(1,\mathrm{IV})}$	$x^{(6,1,1)}_{(2,\mathrm{IV})}$	$x^{(6,1,1)}_{(3,\mathrm{IV})}$	$y^{(6,1,1)}_{\mathrm{IV}}$
1	17.9	1.65	0.87	1+	25	23.15	3.11	2.52	0
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	15.29	1.56	0.96	1+	27	13.7	1.89	2.55	0
4	20.56	2.29	2.56	0	28	15.19	1.75	1.16	0
5	26.04	4.7	5.26	0	29	14.5	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.2	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.2	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	19.75	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	21.66	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 40　欠測パターン (6, 1, 1) に従った擬似完全データ V.

No.	$x_{(1,\mathrm{V})}^{(6,1,1)}$	$x_{(2,\mathrm{V})}^{(6,1,1)}$	$x_{(3,\mathrm{V})}^{(6,1,1)}$	$y_{\mathrm{V}}^{(6,1,1)}$	No.	$x_{(1,\mathrm{V})}^{(6,1,1)}$	$x_{(2,\mathrm{V})}^{(6,1,1)}$	$x_{(3,\mathrm{V})}^{(6,1,1)}$	$y_{\mathrm{V}}^{(6,1,1)}$
1	16.83	1.65	0.87	1+	25	26.45	3.11	2.52	0
2	20.48	2.19	0.48	0	26	12.44	1.83	0.72	0
3	12.39	1.56	0.96	1+	27	13.7	1.89	2.55	0
4	20.56	2.29	2.56	0	28	14.47	1.75	1.16	0
5	26.04	2.44	1.69	0	29	14.5	2.52	1.49	0
6	19.97	3.35	-0.23	1+	30	18.85	1.74	2.2	0
7	33.08	4.05	4.33	1+	31	18.98	2.35	0.72	0
8	15.75	1.12	1.43	0	32	17.2	1.49	1.16	0
9	24.96	2.31	1.86	1+	33	24.77	3.02	2.04	1+
10	21.59	3.75	-0.67	1+	34	15.43	1.15	1.27	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	24.74	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	19.76	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 41　欠測パターン (6, 2, 1) に従った擬似完全データ I.

No.	$x_{(1,\mathrm{I})}^{(6,2,1)}$	$x_{(2,\mathrm{I})}^{(6,2,1)}$	$x_{(3,\mathrm{I})}^{(6,2,1)}$	$y_{\mathrm{I}}^{(6,2,1)}$	No.	$x_{(1,\mathrm{I})}^{(6,2,1)}$	$x_{(2,\mathrm{I})}^{(6,2,1)}$	$x_{(3,\mathrm{I})}^{(6,2,1)}$	$y_{\mathrm{I}}^{(6,2,1)}$
1	20.25	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	16.32	1.56	0.96	1+	27	17.85	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	26.04	2.57	3.78	0	29	18.85	1.74	2.2	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.2	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	0
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	15.69	2.18	0.72	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	24.28	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	23.59	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 42　欠測パターン (6, 2, 1) に従った擬似完全データ II.

No.	$x_{(1,\mathrm{II})}^{(6,2,1)}$	$x_{(2,\mathrm{II})}^{(6,2,1)}$	$x_{(3,\mathrm{II})}^{(6,2,1)}$	$y_{II}^{(6,2,1)}$	No.	$x_{(1,\mathrm{II})}^{(6,2,1)}$	$x_{(2,\mathrm{II})}^{(6,2,1)}$	$x_{(3,\mathrm{II})}^{(6,2,1)}$	$y_{\mathrm{II}}^{(6,2,1)}$
1	15.21	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	14.23	1.56	0.96	1+	27	19.19	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	26.04	2.01	3.86	0	29	18.85	1.74	2.2	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.2	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	0
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	16.09	1.43	0.72	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	26.13	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	23.45	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 43　欠測パターン (6, 2, 1) に従った擬似完全データ III.

No.	$x_{(1,\mathrm{III})}^{(6,2,1)}$	$x_{(2,\mathrm{III})}^{(6,2,1)}$	$x_{(3,\mathrm{III})}^{(6,2,1)}$	$y_{III}^{(6,2,1)}$	No.	$x_{(1,\mathrm{III})}^{(6,2,1)}$	$x_{(2,\mathrm{III})}^{(6,2,1)}$	$x_{(3,\mathrm{III})}^{(6,2,1)}$	$y_{\mathrm{III}}^{(6,2,1)}$
1	11.2	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	20.22	1.56	0.96	1+	27	16.51	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	26.04	2.86	0.59	0	29	18.85	1.74	2.2	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.2	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	0
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	17.07	2.92	0.72	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	21.36	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	23.36	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 44　欠測パターン (6, 2, 1) に従った擬似完全データ IV.

No.	$x_{(1,\mathrm{IV})}^{(6,2,1)}$	$x_{(2,\mathrm{IV})}^{(6,2,1)}$	$x_{(3,\mathrm{IV})}^{(6,2,1)}$	$y_{\mathrm{IV}}^{(6,2,1)}$	No.	$x_{(1,\mathrm{IV})}^{(6,2,1)}$	$x_{(2,\mathrm{IV})}^{(6,2,1)}$	$x_{(3,\mathrm{IV})}^{(6,2,1)}$	$y_{\mathrm{IV}}^{(6,2,1)}$
1	14.74	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	18.5	1.56	0.96	1+	27	16.73	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	26.04	2.64	5.16	0	29	18.85	1.74	2.2	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.2	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	0
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	16.57	2.3	0.72	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	22.15	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	20.52	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 45　欠測パターン (6, 2, 1) に従った擬似完全データ V.

No.	$x_{(1,\mathrm{V})}^{(6,2,1)}$	$x_{(2,\mathrm{V})}^{(6,2,1)}$	$x_{(3,\mathrm{V})}^{(6,2,1)}$	$y_{\mathrm{V}}^{(6,2,1)}$	No.	$x_{(1,\mathrm{V})}^{(6,2,1)}$	$x_{(2,\mathrm{V})}^{(6,2,1)}$	$x_{(3,\mathrm{V})}^{(6,2,1)}$	$y_{\mathrm{V}}^{(6,2,1)}$
1	14.19	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	15.46	1.56	0.96	1+	27	18.98	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	26.04	2.88	1.59	0	29	18.85	1.74	2.2	0
6	19.97	3.35	-0.23	1+	30	18.98	2.35	0.72	0
7	33.08	4.05	4.33	1+	31	17.2	1.49	1.16	0
8	15.75	1.12	1.43	0	32	24.77	3.02	2.04	0
9	24.96	2.31	1.86	1+	33	15.43	1.15	1.27	1+
10	21.59	3.75	-0.67	1+	34	12.79	1.53	0.72	1+
11	27.36	2.3	1.75	0	35	14.53	1.45	0	1+
12	18.86	1.96	1.75	1+	36	15.22	1.39	1.04	0
13	12.12	0.72	3.2	1+	37	12.41	1.15	0.41	0
14	22.83	3.35	0.54	0	38	19.71	1.96	-0.67	0
15	13.32	1.49	-0.1	0	39	20.96	1.75	1.75	0
16	20.69	2.71	0.1	1+	40	16.64	1.76	0.41	0
17	15.75	2.15	-0.67	0	41	16.74	1.6	0	0
18	25.05	2.71	0.72	0	42	21.22	2.74	1.23	1+
19	25.57	3.11	3.68	1+	43	19.53	2.31	2.59	1+
20	15.26	2.46	0.96	1+	44	21.37	3.56	2.73	0
21	27.85	2.13	2.15	1+	45	20.44	3.27	1.94	1+
22	15.26	2.46	0.96	0	46	25.97	2.61	1.37	1+
23	16.89	1.49	0.72	0	47	18.14	2.79	1.75	0
24	13.29	2.71	-2.45	0	48	20.29	1.43	2.82	1+

表 46　欠測パターン (7, 1, 0) に従った擬似完全データ I.

No.	$x_{(1,\mathrm{I})}^{(7,1,0)}$	$x_{(2,\mathrm{I})}^{(7,1,0)}$	$x_{(3,\mathrm{I})}^{(7,1,0)}$	$y_{\mathrm{I}}^{(7,1,0)}$	No.	$x_{(1,\mathrm{I})}^{(7,1,0)}$	$x_{(2,\mathrm{I})}^{(7,1,0)}$	$x_{(3,\mathrm{I})}^{(7,1,0)}$	$y_{\mathrm{I}}^{(7,1,0)}$
1	16.32	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	19.97	1.56	0.96	1+	27	17.79	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.2	0
6	33.08	4.05	4.33	0	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	1+	31	17.2	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	0
9	21.59	3.75	-0.67	0	33	15.43	1.15	1.27	1+
10	27.36	2.3	1.75	1+	34	15.55	2.16	0.72	1+
11	18.86	1.96	1.75	0	35	14.53	1.45	0	1+
12	12.12	0.72	3.2	1+	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	1+	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.1	0	38	19.71	1.96	-0.67	0
15	20.69	2.71	0.1	0	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	1+	40	16.64	1.76	0.41	0
17	25.05	2.71	0.72	0	41	16.74	1.6	0	0
18	25.57	3.11	3.68	0	42	24.14	2.74	1.23	1+
19	15.26	2.46	0.96	1+	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	0
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	1+	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	0	47	23.46	2.79	1.75	0
24	24.48	3.11	2.52	0	48	20.29	1.43	2.82	1+

表 47　欠測パターン (7, 1, 0) に従った擬似完全データ II.

No.	$x_{(1,\mathrm{II})}^{(7,1,0)}$	$x_{(2,\mathrm{II})}^{(7,1,0)}$	$x_{(3,\mathrm{II})}^{(7,1,0)}$	$y_{\mathrm{II}}^{(7,1,0)}$	No.	$x_{(1,\mathrm{II})}^{(7,1,0)}$	$x_{(2,\mathrm{II})}^{(7,1,0)}$	$x_{(3,\mathrm{II})}^{(7,1,0)}$	$y_{\mathrm{II}}^{(7,1,0)}$
1	16.87	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	14.99	1.56	0.96	1+	27	19.19	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.2	0
6	33.08	4.05	4.33	0	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	1+	31	17.2	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	0
9	21.59	3.75	-0.67	0	33	15.43	1.15	1.27	1+
10	27.36	2.3	1.75	1+	34	16.07	1.43	0.72	1+
11	18.86	1.96	1.75	0	35	14.53	1.45	0	1+
12	12.12	0.72	3.2	1+	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	1+	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.1	0	38	19.71	1.96	-0.67	0
15	20.69	2.71	0.1	0	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	1+	40	16.64	1.76	0.41	0
17	25.05	2.71	0.72	0	41	16.74	1.6	0	0
18	25.57	3.11	3.68	0	42	26.13	2.74	1.23	1+
19	15.26	2.46	0.96	1+	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	0
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	1+	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	0	47	23.45	2.79	1.75	0
24	22.67	3.11	2.52	0	48	20.29	1.43	2.82	1+

表 48　欠測パターン (7, 1, 0) に従った擬似完全データ III.

No.	$x_{(1,\,\mathrm{III})}^{(7,\,1,\,0)}$	$x_{(2,\,\mathrm{III})}^{(7,1,0)}$	$x_{(3,\,\mathrm{III})}^{(7,1,0)}$	$y_{\mathrm{III}}^{(7,1,0)}$	No.	$x_{(1,\,\mathrm{III})}^{(7,\,1,\,0)}$	$x_{(2,\,\mathrm{III})}^{(7,1,0)}$	$x_{(3,\,\mathrm{III})}^{(7,1,0)}$	$y_{\mathrm{III}}^{(7,1,0)}$
1	14.13	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	11.06	1.56	0.96	1+	27	16.44	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.2	0
6	33.08	4.05	4.33	0	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	1+	31	17.2	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	0
9	21.59	3.75	-0.67	0	33	15.43	1.15	1.27	1+
10	27.36	2.3	1.75	1+	34	16.9	2.9	0.72	1+
11	18.86	1.96	1.75	0	35	14.53	1.45	0	1+
12	12.12	0.72	3.2	1+	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	1+	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.1	0	38	19.71	1.96	-0.67	0
15	20.69	2.71	0.1	0	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	1+	40	16.64	1.76	0.41	0
17	25.05	2.71	0.72	0	41	16.74	1.6	0	0
18	25.57	3.11	3.68	0	42	21.19	2.74	1.23	1+
19	15.26	2.46	0.96	1+	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	0
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	1+	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	0	47	23.15	2.79	1.75	0
24	27.4	3.11	2.52	0	48	20.29	1.43	2.82	1+

表 49　欠測パターン (7, 1, 0) に従った擬似完全データ IV.

No.	$x_{(1,\,\mathrm{IV})}^{(7,\,1,\,0)}$	$x_{(2,\,\mathrm{IV})}^{(7,1,0)}$	$x_{(3,\,\mathrm{IV})}^{(7,1,0)}$	$y_{\mathrm{IV}}^{(7,1,0)}$	No.	$x_{(1,\,\mathrm{IV})}^{(7,\,1,\,0)}$	$x_{(2,\,\mathrm{IV})}^{(7,1,0)}$	$x_{(3,\,\mathrm{IV})}^{(7,1,0)}$	$y_{\mathrm{IV}}^{(7,1,0)}$
1	14.6	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	14.53	1.56	0.96	1+	27	16.66	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.2	0
6	33.08	4.05	4.33	0	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	1+	31	17.2	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	0
9	21.59	3.75	-0.67	0	33	15.43	1.15	1.27	1+
10	27.36	2.3	1.75	1+	34	16.33	2.25	0.72	1+
11	18.86	1.96	1.75	0	35	14.53	1.45	0	1+
12	12.12	0.72	3.2	1+	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	1+	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.1	0	38	19.71	1.96	-0.67	0
15	20.69	2.71	0.1	0	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	1+	40	16.64	1.76	0.41	0
17	25.05	2.71	0.72	0	41	16.74	1.6	0	0
18	25.57	3.11	3.68	0	42	21.91	2.74	1.23	1+
19	15.26	2.46	0.96	1+	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	0
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	1+	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	0	47	20.3	2.79	1.75	0
24	25.45	3.11	2.52	0	48	20.29	1.43	2.82	1+

表 50　欠測パターン (7, 1, 0) に従った擬似完全データ V.

No.	$x_{(1,\mathrm{V})}^{(7,1,0)}$	$x_{(2,\mathrm{V})}^{(7,1,0)}$	$x_{(3,\mathrm{V})}^{(7,1,0)}$	$y_{\mathrm{V}}^{(7,1,0)}$	No.	$x_{(1,\mathrm{V})}^{(7,1,0)}$	$x_{(2,\mathrm{V})}^{(7,1,0)}$	$x_{(3,\mathrm{V})}^{(7,1,0)}$	$y_{\mathrm{V}}^{(7,1,0)}$
1	22	1.65	0.87	1+	25	12.44	1.83	0.72	0
2	20.48	2.19	0.48	0	26	13.7	1.89	2.55	0
3	14.1	1.56	0.96	1+	27	18.98	1.75	1.16	0
4	20.56	2.29	2.56	0	28	14.5	2.52	1.49	0
5	19.97	3.35	-0.23	1+	29	18.85	1.74	2.2	0
6	33.08	4.05	4.33	0	30	18.98	2.35	0.72	0
7	15.75	1.12	1.43	1+	31	17.2	1.49	1.16	0
8	24.96	2.31	1.86	1+	32	24.77	3.02	2.04	0
9	21.59	3.75	-0.67	0	33	15.43	1.15	1.27	1+
10	27.36	2.3	1.75	1+	34	12.9	1.56	0.72	1+
11	18.86	1.96	1.75	0	35	14.53	1.45	0	1+
12	12.12	0.72	3.2	1+	36	15.22	1.39	1.04	0
13	22.83	3.35	0.54	1+	37	12.41	1.15	0.41	0
14	13.32	1.49	-0.1	0	38	19.71	1.96	-0.67	0
15	20.69	2.71	0.1	0	39	20.96	1.75	1.75	0
16	15.75	2.15	-0.67	1+	40	16.64	1.76	0.41	0
17	25.05	2.71	0.72	0	41	16.74	1.6	0	0
18	25.57	3.11	3.68	0	42	21.22	2.74	1.23	1+
19	15.26	2.46	0.96	1+	43	19.53	2.31	2.59	1+
20	27.85	2.13	2.15	1+	44	21.37	3.56	2.73	0
21	15.26	2.46	0.96	0	45	20.44	3.27	1.94	1+
22	16.89	1.49	0.72	1+	46	25.97	2.61	1.37	1+
23	13.29	2.71	-2.45	0	47	18.14	2.79	1.75	0
24	23.98	3.11	2.52	0	48	20.29	1.43	2.82	1+